

Oct 12 1952

X

Psychological Bulletin

SCIENCE
LIBRARY

CONTENTS

ARTICLES:

Psychological Tests in Psychopathological Prognosis: CHARLES WINDLE, 451

Some Observations on Q Technique: WILLIAM STEPHENSON, 483

The Three Basic Factor-Analytic Research Designs—Their Interrelations and Derivatives: RAYMOND B. CATTELL, 499

Some Greco-Latin Analysis of Variance Designs for Learning Studies: E. JAMES ARCHER, 521

NOTES:

A Note on Profile Similarity: HAROLD WEBSTER, 538

On the Reliability of the Leaderless Group Discussion Technique: BORIS SEMEONOFF, 540

Genetics of Schizophrenia: A Rejoinder: NICHOLAS PASTORE, 542

The Genetics of Schizophrenia: Further Rejoinder to Pastore: LEWIS A. HURST, 544

BOOK REVIEWS:

HUMPHREY'S Thinking: An Introduction to its experimental psychology: GEORGE A. MILLER, 547

KEMPTHORNE'S The design and analysis of experiments: C. J. BURKE, 548

JAHODA, DEUTSCH, AND COOK'S Research: methods in social relations: With especial reference to prejudice: J. C. GILCHRIST, 549

TILTON'S An educational psychology of learning: J. B. SPROUD, 551

FILM REVIEWS:

PRONKO AND SNYDER'S Vision with spatial inversion: DOROTHEA E. JOHANNSEN, 553

MILLER AND HUNT'S Motivation and reward in learning: TRACY S. KENDLER, 554

HAYES AND HAYES' Vocalization and speech in chimpanzees: EDWARD M. BENNETT, 555

STONE'S When should grown-ups help? and And then ice cream (children's meals): ELIAS KATZ, 556

VIERLANCE'S Testing intelligence with the Stanford-Binet: ELIAS KATZ, 557

BINNET AND MCKEEVER'S "Antabuse" in the treatment of alcoholism: ALBERT D. ULLMAN, 557

PUBLISHED BIMONTHLY BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

LYLE H. LANIER, Editor
University of Illinois

LORRAINE BOUTHILLET, Managing Editor

CONSULTING EDITORS

STUART H. BRITT
Northam, Louis and Brerley, Inc., Chicago

DORWIN CARTWRIGHT
University of Michigan

FRANK A. GELDARD
University of Virginia

JAMES J. GIBSON
Cornell University

DAVID A. GRANT
University of Wisconsin

WILLIAM T. HERON
University of Minnesota

ERNEST R. HILGAARD
Stanford University

WILLIAM A. HUNT
Northwestern University

JEAN WALKER McFARLANE
University of California

DONALD G. MILLER
University of Michigan

JOHN T. MURPHY
University of Vermont

JAMES G. MILLER
University of Chicago

NEAL E. MILLER
Yale University

HELEN F. MILLER
University of Michigan

ROBERT H. MURPHY
Harvard University

ROBERT L. MURPHY
Teachers College, Columbia University

The *Psychological Bulletin* contains evaluative summaries of the literature in various fields of psychology, methodological articles, and book reviews. This JOURNAL does not publish reports of original research or original theoretical articles.

Editorial communications concerning items to be published during 1952 should be sent to Lyle H. Lanier, Department of Psychology, University of Illinois, Urbana, Illinois. All new manuscripts should henceforth be sent to the incoming editor, Wayne Denuis, Department of Psychology, Brooklyn College, Brooklyn, New York. Books for review should be sent to the new book review editor, Edward Girden, at the same address.

Preparation of articles for publication. Authors are strongly advised to follow the general directions given in the "Publication Manual of the American Psychological Association," by the Council of Editors of the American Psychological Association (*Psychological Bulletin*, 1952, 49, (No. 1, Supplement), 389-446). Special attention should be given to the section on the preparation of the references (pp. 432-440), since this is a particular source of difficulty in long reviews of research literature. All copy must be double spaced, including the references.

Reprints. Authors may order reprints when returning proof. Five copies of the JOURNAL are supplied gratis to contributors of articles, notes, special reviews, and book reviews. No reprints are supplied gratis.

Business communications—including subscription orders of back issues and changes of address—should be sent to the American Psychological Association, 1515 Massachusetts Avenue, N. W., Washington, D. C.

Annual subscription: \$7.00 (Foreign \$7.50). Single copies, \$1.25.

PUBLISHED BIMONTHLY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

1515 Massachusetts Ave. N.W., Washington, D. C.

Entered as second class mail matter at the post office at Washington, D. C., under the act of March 3, 1879. Additional entry at the post office at Menasha, Wisconsin. Accepted for mailing at special rate of postage under the provisions of Sec. 34-40 Par. (2) provided for in Section 1103 of Act of February 23, 1949, authorized August 6, 1947. Printed in U.S.A.

Copyright, 1952, by The American Psychological Association, Inc.

Psychological Bulletin

PSYCHOLOGICAL TESTS IN PSYCHOPATHOLOGICAL PROGNOSIS¹

CHARLES WINDLE

University of Iowa

Prognosis in psychopathology is essentially a prediction specifying a relationship between characteristics of a psychopathological condition and the eventual outcome of the disorder. In the field of psychological tests, the prognosis will relate certain aspects of test performance (such as the number of Rorschach *F*+ responses or the Wechsler-Bellevue IQ) to the degree of improvement in mental health of the patients in question.

Although prognostic information is of great practical as well as theoretical value, investigative work in this field appears discouragingly unorganized. It is hoped that a critical review of that aspect of the problem of most interest to psychologists, viz., the prognostic use of psychological tests, may help investigators plan more fruitful research. Previous reviews (10, 42, 44, 77, 103, 113, 136, 164) of this topic have tended to be predominantly descriptive and of limited scope. There remains a need for a comprehensive account of the evidence with sufficiently critical evaluation of it.

Since relatively little work in this field has involved cross validation, evaluation in this review will be based primarily upon comparisons among findings. The material has been organized on the basis of the particular psychological test employed. Although a classification based on either diagnostic category or type of therapy might have been more desirable, such a classification seemed impossible in view of the data available in the studies to be reviewed. Very frequently the patients employed in a single study covered a range of diagnostic categories—either implicitly as when “psychotics” were studied, or explicitly as when two subcategories of schizophrenia were listed and then considered together because no diagnostic differences in prognosis appeared. Furthermore, the meaning of diagnostic labels is often uncertain. A similar

¹ The writer is greatly indebted to Dr. Joseph Zubin for his guidance and encouragement in the preparation of this paper.

confusion exists in respect to type of therapy; results from different therapeutic procedures have often been grouped together and in some cases none has been specified. Consequently, studies have been classified in terms of the psychological tests used. But within each category of psychological test, studies have been classified as far as possible in terms of varieties of mental disorder. Insofar as these diagnostic categories are meaningful, it is possible for the reader to relate prognostic test results to them. As a source of reference and basis for comparisons a master table (Table 1) has been compiled in which some of the most important characteristics of the studies in this field which the author judged to be the best are presented. The text of the article will attempt to elaborate the most important of the specific criticisms implicit in the table and to evaluate the present use of specific psychological tests for prognosis.

The term *prognosis* is used to cover a large range of predictive relationships. The meaning of prognosis is a function of three kinds of variables: (a) the situation from which the prediction is made, (b) the intervening conditions which may influence the eventual outcome, and (c) the final conditions predicted.

Within the first category fall such factors as the nature, duration and severity of the disease, and the patient's age, sex, attitudes, and capacities. Psychological test performance would be included in this category. To test the prognostic efficacy of any one factor it is necessary to hold the others constant, an ideal which usually can only be approximated. Obviously, the more thorough the experimental and statistical control of confounding variables, the greater the value of the study.

Among the more important intervening conditions are the type of therapy employed, if any, and the amount of time elapsing prior to evaluation of outcome. The base line for this evaluation under different therapies is the rate of the so-called "spontaneous remission," or the expected rate of improvement when no specific therapy is applied. It has been reported that regardless of the type of therapy employed, little, if any, enhancement above the rate of spontaneous remissions is gained (34, 173). It seems that at the present state of therapeutic development some characteristic of the individual is the important factor, not the specific therapy. Perhaps if each individual were to be given the type of therapy most suitable to his particular endowment, the rate of improvement might be enhanced. When the prognostic criteria for spontaneous remissions are established, it will be easier to develop the prognostic criteria for specific therapies.

Experimenter	N	Type of Patients	Duration of Illness	Follow-up Time	Therapy	Prognostic Index	Prognosis	Statistical Significance
Graham (57)	18	Sc	3 yr. (Md)	Im	I ME	Many characteristic responses		$\bar{D} < G^*$
McCall (102)	27	Pa	L	34 yr.	BS or TP	High <i>Plims</i> (plant importance) High <i>P</i> (popularity of response)	F U	$\bar{F} < .02$ $\bar{U} < .05$
Katz (80)	113	Sc	NG	Im	I ECT	Long CRT (reaction time to colored cards) Long TRT (reaction time to all cards) High <i>DeH</i> (dehumanization) Ascendance attributed to perceived figures High <i>SaC</i> (surface color) No Rorschach prognostic index Single Rorschach factors not prognostic No Rorschach prognostic index	F F U F F	$\bar{F} < .05$ $\bar{F} < .05$ $\bar{U} < .05$ $\bar{F} < .05$ $\bar{F} < .05$
Rabin (130)	10	Pa	Sh	Im & 6-18 mo.	I	Piotrowski's 1941 criterion (≥ 3 of 6 signs) (126)	not prognostic	$\bar{F} < .5$
Wechsler <i>et al.</i> (169)	15	Sc	NG	Im	Py	Piotrowski's 1941 criterion (≥ 3 of 6 signs) (126)	not prognostic	$\bar{F} < .7$
Rioch (134)	36	Pp	NG	Im & ≥ 3 yr.	I	Piotrowski's 1941 criterion (≥ 3 of 6 signs) (126)	not prognostic	$\bar{F} < .8$
Rees & Jones (131)	35	Sc	NG	Im & ≥ 3 yr.	ECT	Piotrowski's 1941 criterion (≥ 3 of 6 signs) (126)	not prognostic	$\bar{F} < .7$
	19	Sc	NG	Im & ≥ 3 yr.	BS	Piotrowski's 1941 criterion (≥ 3 of 6 signs) (126)	not prognostic	$\bar{F} < .3$
	22	Sc	NG	Im & ≥ 3 yr.	BS	Of > 20 signs, none prognostic ¹	not prognostic	$\bar{F} < .3$
	10	Sc	NG	Im & ≥ 3 yr.	BS	Multiple-choice Rorschach indices	not prognostic	$\bar{F} < .05$
Scherer (141)	86	Sc	V	2-4 wks.	BECT	Low rejection	not prognostic	$\bar{F} < .05$
Scherer (142)	41	Sc & MD	NG	2-4 wks.	No	High <i>M</i>	not prognostic	$\bar{F} < .001$
	26	Pp ^a	NG	NG	BS	Multiple-choice Rorschach indices	not prognostic	$\bar{F} < .001$
	22	Pp	NG	NG	No	Of > 50 signs only high <i>R</i> to cards VIII, IX, & X	not prognostic	$\bar{F} < .05$
Filmer-Bennett (44)	22	Pp ^a	NG	NG	No	Five psychodynamic cluster patterns	not prognostic	$\bar{F} < .05$
	24	Pp	NG	> 1 yr.	V	Seven patterns of signs	not prognostic	$\bar{F} < .05$
	54	Pp	NG	> 1 yr.	NG	Combination of the best of the above patterns of signs	not prognostic	$\bar{F} < .02$ (each)
						Individual consideration of 7 pattern variables	not prognostic	$\bar{F} < .02$ (each)
Roberts (135)	78	Pp	NG	> 1 yr.	V	Basic Rorschach Score of Euhler, Buhler, & Lefever	not prognostic	SS (each)
Siegel (154)	51	N (?)	NG	Im	Py	Eleven indices derived from the literature	not prognostic	SS (each)
	96 ^a	PaN	NG	Im	(PaPy)	Of many separate signs only 3 (<i>FC</i> , <i>Aobj</i> , & <i>R</i>) prognostic	not prognostic	$\bar{F} < .05$ (each)
						Of several trait syndromes only 3 prognostic	not prognostic	$\bar{F} < .05$ (each)
						Five blind predictions	not prognostic	$\bar{F} < .05$ (each)
Dickson (36)	35	PaN	NG	Im	(Py)	High <i>H%</i> High <i>P</i> , presence of <i>Fc</i> Presence of <i>M</i> , and 3 or more <i>M</i> $F\% > 80$, presence of <i>CF</i> $F\% > (C/F+C)$	$\bar{F} < .02$ (each) $\bar{F} < .05$ (each) $\bar{F} < .01$ (each) $\bar{F} < .05$ (each) $\bar{F} < .10$ (each)	$\bar{F} < .02$ (each) $\bar{F} < .05$ (each) $\bar{F} < .01$ (each) $\bar{F} < .05$ (each) $\bar{F} < .10$ (each)
Bradway <i>et al.</i> (18)	16	P+G	NG	EP	Ps & CW	$PM \geq 2$; FK ; Fc ; mean response time < 40 sec; rejected cards; and $\%R$ to last 3 cards ≥ 40 8 or more of these 10 signs	$\bar{F} < .02$ (each) $\bar{F} < .05$ (each) $\bar{F} < .01$ (each) $\bar{F} < .05$ (each) $\bar{F} < .10$ (each)	$\bar{F} < .02$ (each) $\bar{F} < .05$ (each) $\bar{F} < .01$ (each) $\bar{F} < .05$ (each) $\bar{F} < .10$ (each)
						5 or more of these 10 signs	$\bar{F} < .02$ (each) $\bar{F} < .05$ (each) $\bar{F} < .01$ (each) $\bar{F} < .05$ (each) $\bar{F} < .10$ (each)	$\bar{F} < .02$ (each) $\bar{F} < .05$ (each) $\bar{F} < .01$ (each) $\bar{F} < .05$ (each) $\bar{F} < .10$ (each)
Krugman (90)	NG	PC	NG	NG	NG	No single factor can determine treatability; 19 factors or aspects listed as prognostic	$\bar{F} < .02$ (each) $\bar{F} < .05$ (each) $\bar{F} < .01$ (each) $\bar{F} < .05$ (each) $\bar{F} < .10$ (each)	$\bar{F} < .02$ (each) $\bar{F} < .05$ (each) $\bar{F} < .01$ (each) $\bar{F} < .05$ (each) $\bar{F} < .10$ (each)
Siegel (153)	26	PC	NG	1 yr.	Py	Refusals, <i>FC</i> , <i>W%</i> , <i>Fc</i> , <i>H</i> , <i>O</i> , <i>P</i> +, and favorable response to Testing the limits	$\bar{F} < .02$ (each) $\bar{F} < .05$ (each) $\bar{F} < .01$ (each) $\bar{F} < .05$ (each) $\bar{F} < .10$ (each)	$\bar{F} < .02$ (each) $\bar{F} < .05$ (each) $\bar{F} < .01$ (each) $\bar{F} < .05$ (each) $\bar{F} < .10$ (each)

¹ The authors concluded that "a number of features showed statistically significant differences between the recovered and not improved groups, but none of these showed a statistically significant predictive value" (131, p. 689). This conclusion may be subject to disagreement.

² Chosen to match the group above.

³ Subdivided to form original and cross-validating groups; these indices applied to both groups.

TABLE 1, continued

Experimenter	N	Type of Patients	Duration of Illness	Follow-up Time	Therapy	Prognostic Index	Prognosis	Statistical Significance
Billig & Sullivan (13) ^a	29	A	NG	1 yr.	NG	O-, CF-, Shading Shock, and CF High score: $2\left(\frac{W}{W+}\leq 33\%; D<50\%; D4+S\geq 10\%; F\geq 60\%; 1(h>1; K>1; e>1; M<FM; FC\leq 1; FC<(CF+C); VIII+IX+X<30\%; Hd>H; p^1>1; R_{EX}$	U	D-NG
	11			none				p-NG
Sillman (155)	42	A	NG	2 yr.	NG	$AI>1; W<M$ Rigidity in responses; blocked efforts after emotional gratification; low capacity for extraversion, verbalization, identification and love; more anxiety, internalized aggression and sense of inferiority	U	D-G
Sloan (159)	30	MtD	-	SM	WP	Beck's criteria of good adjustment: $(P+\geq 50\%; P=4 \text{ or } 5; \text{ no } C; CF=1 \text{ or } 2; C \text{ sum } <3; A\%$ 50-80; $M<2$ Individual quantitative indices Recoverability from color and shading shock	U not prognostic F	p-NG D-NG p-NG D-G
Harris & Christiansen (67)	35	DRPD	-	Im	EPy(Pa)	Location, determinant and content scores Empirical weighted score: $(+2(M, FM, Anat-sex; +1.5(PK); +1(Pe, C, C); -5(P); -1(K, F-, Rq); -2(C, FC); -2.5(CF)+4 \text{ or more}$ Subjective evaluation (movement and no color) Many anatomy responses (60% or more) Normal pattern, neurotic pattern with guilt feelings High R, high Z High M, high sum C, high P, many different content categories Rejection of card	F F U F F F F F U	p<.05 p<.05 D-NG D-NG p=.01 (each) p=.05 (each) D-NG D-G D-NG D-NG
Levi (93, 94)	20	PH	NG	NG	PR			
Ellis & Brown (39)	26	PT	NG	CT	-			
Rabin (129)	1	S & M	-	2 mo.	-			
Davidson & Conkey (32)	20	GP	NG	1 yr.	Ma	Absence of emotional expression and evidence of psychic tension (color and shading shock) Piotrowski's signs of organic disorder	U U	p-NG D-NG
Free Drawing Test								
Fiedler & Siegel (43)	34	PaN	NG	NG	Py	Goodenough's criteria of drawing the head are met	F	p=.03
Minnesota Multiphasic Personality Inventory								
Harris (64)	NG	Pp	NG	NG	S	High scores in psychasthenia and psychotic scales	U	p-NG D-NG
Harris et al. (66)	31	Pp	L	Im	Ea	High scores on only D scale or on several scales including D, Sc, Pa, and Pa Normal based on criteria Sortal, based on criteria *developed in other	F U	p-NG D-NG p-NG D-NG
Experimenter	N	Type of Patients	Duration of Illness	Follow-up Time	Therapy	Prognostic Index	Prognosis	Statistical Significance

Experimenter	N	Type of Patients	Duration of Illness	Follow-up Time	Therapy	Prognostic Index	Prognosis	Statistical Significance
Hales & Simon (60)	20	Sc	3 yr. 2 mo. (M)	Im	I	MMPI scores and profiles not prognostic High scores in psychotic components	F	\bar{p} -NG D-NG
Pearson (119)	29	Sc	NG	Im	ECT	Score on D scale >84 or <55 ; Score on S ₂ scale >84 ; Scores ≥ 80 on 4 or more scales	U	$\bar{p} < .01$ (?) \bar{p} -NG D-G***
Pearson & Swenson (120)	33	Pp	NG	Im	ECT	Validation of the above empirical finding		$\bar{p} < .01$ (each)
Feldman (41)	40	Pp	NG	1 yr.	S	42 MMPI items differential of outcome		$\bar{p} < .01$
Feldman (42)	44	Pp	NG	NG	NG	Validation of this scale		$\bar{p} < .01$
	84	Pp	NG	3, 6, & 12 mo.	S	High P ₂ , P ₁ , S ₂ , & M ₂ items; unfavorable in F & K	U	SS (each) $\bar{p} = .0002$ $\bar{p} = .02-.01$ $\bar{p} = .001$
Pacella et al. (118)	100	Pp	NG	6 mo.	S	Validation of P ₂ scale differential of outcome	not prognostic	\bar{p} -NG
	33	Pp	NG	Im	BPp	Abbreviated P ₂ scale differential of outcome		$\bar{p} < .05$ (each)
	100	Pp	NG	6 mo.	ECT	Blind sorting of MMPI records for prognosis		$\bar{p} < .05$ (each)
	40	Sc	NG	1 mo.	ECT	Pre sorting MMPI		$\bar{p} < .05$ (each)
Carp (20)	37	Sc	13.7 mo. (Md) 11.4 mo. (σ)	NG	I	Low P ₂ , P ₁ , and S ₂ scores; high K scores; low (S ₂ +K) scores		$\bar{p} < .05$ (each)
Schofield (146)	48	Pp	NG	NG	NG	Therapy-susceptible items (P ₂ scale)	U	$\bar{p} < .05$ (each)
Harris & Christiansen (67)	53	DRP	—	Im	BPp (Pa)	High scores on P ₂ , P ₁ , P ₂ , S ₂ , & M ₂ scales	not prognostic	$\bar{p} = .20$ or $< .02$ $\bar{p} < .05$ (each)
Ruesch et al. (138)	20	NRT	—	NG	—	Blind classification by clinical criteria		$\bar{p} < .05$ (each)
	18	DRT	—	NG	—	8 of 35 empirically derived subscales		$\bar{p} < .05$ (each)
						Delayed recovery cases have high scores on P ₂ , P ₁ , and neurotic scales		$\bar{p} < .02$
Albee (3)	127	Pp	NG	>1 yr.	NG	Measures of Aggression	U	$\bar{p} < .001$
Albee & Goldman (5)	81	Sc	NG	>1 yr.	NG	Extrapunitive behavior	U	$\bar{p} < .001$
	44	Pp	NG	NG	NG	Picture-Frustration E% and I%	not prognostic	
Langfeldt (92)	83	Sc	NG	7 yr.	NG	Ability Tests	U	\bar{p} -NG D-G**
Terry & Rennie (163)	47	Sc	NG	4-5 yr.	NG	Ratings of "clever" or "intellectually debilitated"	not prognostic	\bar{p} -NG D-G***
Malumud & Rander (105)	177	Sc	NG	5-5.5 yr.	NG	Intelligence Binet IQ above 105	U	\bar{p} -NG D-G***
Graham (57)	309	Sc	3 yr. (Md)	V	I & M ₂	Binet "scatter" of 5 years or more	U	\bar{p} -NG D-G***
O'Connell & Penrose (116)	30	Pp	NG	NG	M ₂	Malumud-Palmer index on Binet; Binet scatter; vocabulary test; digits forward & backward; free association; arithmetic; interpretation of fables	not prognostic	\bar{p} -NG D-G**
Schnack et al. (143)	31	Sc	NG	V	I or M ₂	Detection of verbal absurdities	F	\bar{p} -NG D-NG
	23-30	Sc	6.1 mo. (M) 16.5 (σ)	Sev	I or I+M ₂	Incompetence in reaction time to auditory stimuli, in tapping rate and in strength of grip	not prognostic	\bar{p} -NG D-NG
						A set of Binet, Word Association and aspiration test signs		\bar{p} -NG D-G
						The above set of signs		
						High scores in 4 of 13 signs from the above tests	F	$\bar{p} < .10$

High scores on only D scale or on several scales in-
cluded in P₂, S₂, P₁, and P₂
Normal scores on all scales
Sorting based on criteria *developed in other
studies

Albee (3)

TABLE 1, continued

Experimenter	N	Type of Patients	Duration of Illness	Follow-up Time	Therapy	Prognostic Index	Prognosis	Statistical Significance
Louie et al. (98)	14-22	Sc	7.2 mo. (M) 16.5 (s)	Sev	Mz	More inulin (the above 13) than metrazol (the below 11) signs Low scores in 6 of 11 different signs	F	$p=.05$ $p<.10$
	20	CSc	10 mo. (M)	4-11 yr.	V	More metrazol than inulin signs Stanford-Binet <70 or child mute	F U	$p=.20$ $p<.05$ D-NG D-G
Kriegman & Hilgard (99)	30	PC	NG	≥ 7 mo.	Py	Higher performance scores (Merrill-Palmer, Arthur, and Pintner-Patterson) than verbal (Stanford-Binet)	F	D-NG D-G***
Rabin (100)	10	Pa	NG	NG	ECT	Stanford-Binet IQ ≥ 110	F	$p<.01$
Peters (121)	71	Pa	L	>1 yr.	BS	Wechsler-Bellevue IQ and memory scales	not prognostic	
Harris et al. (66)	24	Pp	L	Im	En	Ratings of normal intelligence Abbreviated Wechsler-Bellevue IQ Higher scores on untimed than timed tests	not prognostic	
Malamud et al. (106)	47	IM	1 yr.	SM	Ms or Py	Estimated intelligence dull normal or lower	F	D-NG D-G***
Malamud et al. (107)	44	IM	1 yr.	7 yr.	Ms or Py	Estimated intelligence below average	F	D-NG D-G***
Fisher (48)	16 43-8	DpPp DpPp	NG NG	Im Im	ECT ECT	Attitude of feeling bad (one of two scales used) Wechsler-Bellevue IQ, Word Naming, Counting by 3's, Word Association, and Level of Aspiration	F	D-NG D-G***
Scherer (141, 142)	38-41	Sc & MD	V	2-4 wk.	BECT	Of >17 tests only lower scores on delayed than immediate memory Efficiency on Minnesota Rate of Manipulation Test Efficiency on Capps' Categorization Test Of >17 tests only high visual memory Of >17 tests ten tests of memory and abstraction	not prognostic U	$p=.05$ $p=.01$ $p=.01$ $p<.05$
Willenson (172)	21-25 21-22	Pp ^a Pp	NG NG	2-4 wk. NG	No BS	Of >17 tests High Wechsler-Bellevue IQ High Similarities, Picture Arrangement, Block Design, and Object Assembly scores	none prognostic F	$p<.05$ (each) $p>.05$ (each) $p<.01$
	22 110	Pp ^a Pp	NG NG	NG NG	No BS or TP	Of >17 tests High Wechsler-Bellevue IQ Wechsler-Bellevue IQ	not prognostic F	$p<.05$ (each) $p<.01$
Carp (21)	9 42	PNSc Sc	NG 19.2 mo. (M) 12.8 mo. (s)	3 mo. 3 yr.	BS I	High Wechsler-Bellevue IQ	U	$p<.01$
Piotrowski & Lewis (128)	91	Sc	Sh	3 yr.	V	IQ below 90	U	D-NG***
Zubin & Windle (180)	16 35	Pa Pa	L L	2 yr. 2 yr.	BS or TP BS or TP	Good performance on Metonym Test Weighted score derived from criterion groups	U F	$p<.05$ $p<.05$
Landis & Erick (91)	16	Pa	L	2 yr.	BS or TP	"Confusion errors" in the Porteus Maze Test	F	D-NG D-NG
Windle & Hamwi (173)	28 26	Pa Pa	>1 yr. <1 yr.	1 & 3 1/2 yr. 0-9 mo.	BS or TP Py or S	Complex Reaction Time Test score >30 Complex Reaction Time Test score >30 Divergence between these two trends	U F	$p<.05$ $p<.05$
Malamud & Gottlieb (104)	80 64	H MN	NG NG	NG NG	Py Py	Binet IQ above 90 Binet IQ above 90	F F	$p<.10$ $p<.20$
Klugman (86)	104	PAN	NG	Im	CF	High Army General Classification Test score	F	D-NG D-NG
Miles et al. (110)	33	N	> 5 yr. (M)	>2 yr.	Py	High Wechsler-Bellevue IQ	F	$p<.01$

TABLE 1, continued

Experimenter	N	Type of Patients	Duration of Illness	Follow-up Time	Therapy	Prognostic Index	Statistical Significance
--------------	---	------------------	---------------------	----------------	---------	------------------	--------------------------

TABLE 1, continued

Experimenter	N	Type of Patients	Duration of Illness	Follow-up Time	Therapy	Prognostic Index	Prognosis	Statistical Significance
Dickson (36)	67	PaN	NG	Im	Py	IQ	not prognostic	$p < .05$
Ruesch et al. (138)	23	NRT	—	NG	—	High IQ	not prognostic	
Harris & Christiansen (97)	20	DET	—	Im	BPY (Pa)	Abbreviated Wechsler-Bellevue IQ		
	53	DRPD	—			mental age on non-verbal tests $\times 100$		
Hamlin (62)	101	MtD	—	NG	NG	High		$\begin{matrix} \text{F} \\ \text{D} \\ \text{G} \end{matrix}$ NG
				NG	NG	The above pattern score $> 108 < 124$		$\begin{matrix} \text{F} \\ \text{D} \\ \text{G} \end{matrix}$ NG
	48	MtD	—	NG	—	The above pattern score < 80		$\begin{matrix} \text{F} \\ \text{D} \\ \text{G} \end{matrix}$ NG
Earl (38)	66	MtD	—	NG	—	Psychographs on 2 verbal and 2 performance tests: unbiased or biased in favor of performance tests irregular or biased in favor of verbal tests		$\begin{matrix} \text{F} \\ \text{D} \\ \text{G} \end{matrix}$ NG
								$\begin{matrix} \text{F} \\ \text{D} \\ \text{G} \end{matrix}$ NG
<i>Abstraction Tests</i>								
Bolles et al. (16)	19	Sc	27 mo. (Md)	Im	I	Good performance on the Vigotaki, Weigl and BRL sorting tests		$\begin{matrix} \text{F} \\ \text{D} \\ \text{G} \end{matrix}$ NG
Graham (57)	7	Sc	3 yr. (Md)	Im	I & Mz	Bolles battery	not prognostic	SS
Zubin & Thompson (179)	49	Sc	Sh	Im & 6 mo.	I	Good performance on Bolles battery	not prognostic	
	43	Sc	L	Im & 6 mo.	Mz or I + Mz	Bolles battery		
<i>Miscellaneous Tests</i>								
Low et al. (99)	102	Sc	NG	4.5 yr.	NG	Appropriate verbal response to Life Situation Test		$\begin{matrix} \text{F} \\ \text{D} \\ \text{G} \end{matrix}$ NG
	54	MD	NG	4.5 yr.	NG	Appropriate verbal response to Life Situation Test		$\begin{matrix} \text{F} \\ \text{D} \\ \text{G} \end{matrix}$ NG
Adler (2)	47	OP	NG	4.5 yr.	NG	Life Situation Test	not prognostic	$\begin{matrix} \text{F} \\ \text{D} \\ \text{G} \end{matrix}$ NG
	50	Sc	NG	—	NG	Functional matching of pictures correlated with good prognosis ratings		$\begin{matrix} \text{F} \\ \text{D} \\ \text{G} \end{matrix}$ NG
Cattell (23)	64	PC	NG	> 1 yr.	NG	Low score on perseveration tests		$\begin{matrix} \text{F} \\ \text{D} \\ \text{G} \end{matrix}$ NG
	57	PC	NG	> 1 yr.	NG	Extreme scores in fluency tests		$\begin{matrix} \text{F} \\ \text{D} \\ \text{G} \end{matrix}$ NG
	24	PC	NG	> 1 yr.	NG	For "nervous children," low "p" scores		$\begin{matrix} \text{F} \\ \text{D} \\ \text{G} \end{matrix}$ NG
Voth (167)	209	Sc	NG	NG	Mz or No	Extensive or no autokinetic movement		$\begin{matrix} \text{F} \\ \text{D} \\ \text{G} \end{matrix}$ NG
	259	MD	NG	NG	Mz or No	Many directional changes		$\begin{matrix} \text{F} \\ \text{D} \\ \text{G} \end{matrix}$ NG
	43	GP	NG	NG	NG	Extensive or no autokinetic movement		$\begin{matrix} \text{F} \\ \text{D} \\ \text{G} \end{matrix}$ NG
	71	N	NG	NG	NG	No autokinetic movement		$\begin{matrix} \text{F} \\ \text{D} \\ \text{G} \end{matrix}$ NG
	754	Pp	NG	NG	NG	Autokinetic movement		$\begin{matrix} \text{F} \\ \text{D} \\ \text{G} \end{matrix}$ NG
Ballinsky (8)	21	DisSc	11 mo. (Md)	1.5 yr. (M)	—	Extensive or no autokinetic movement	not prognostic	$\begin{matrix} \text{F} \\ \text{D} \\ \text{G} \end{matrix}$ NG
						Positive attitude toward status in society, high drive, emotional expressiveness, small discrepancy between vocabulary & dexterity-spatial tests, and undepressed motility		$\begin{matrix} \text{F} \\ \text{D} \\ \text{G} \end{matrix}$ NG
Peters (121)	71	Ps	L	> 1 yr.	BS	Ratings of + for mood swings and hallucinations; — for energy, purposive, active, initiative, evident interests, and reality grasp		$\begin{matrix} \text{F} \\ \text{D} \\ \text{G} \end{matrix}$ NG
						Ratings of + for restless, aggressive, agitation, and speed or flight of ideas		$\begin{matrix} \text{F} \\ \text{D} \\ \text{G} \end{matrix}$ NG
Harris et al. (63)	14	AC	1-9 yr.	NG	AT	Crying easily and confidence in mothers	not prognostic	$\begin{matrix} \text{F} \\ \text{D} \\ \text{G} \end{matrix}$ NG

Estimates of final outcome of mental diseases may differ when different criteria of improvement are adopted. An individual rated as improved on the basis of socioeconomic status may be considered unimproved on the basis of psychosomatic complaints. In some cases emphasis is placed on the objectivity or reliability of the criteria and indices such as the hospitalized status of the patient or a Distress-Relief Quotient are used. More frequently the emphasis on validity leads to the rejection of such mediated criteria, and clinical opinion serves as the accepted index. Clinical opinion, however, fails to lend itself to accurate description or reproduction, and thus is of dubious reliability among investigations. As yet the problem of optimal measures of adjustment remains unsolved. It is necessary to bear in mind that unless measures of outcome are highly correlated, the meanings of prognoses in different studies will differ.

RORSCHACH INK BLOT TEST

The psychological test which has been used most often in the search for prognostic indices is the Rorschach. According to Hertz, "It is in its prognostic power, perhaps, that the Rorschach has its greatest possibilities" (70, p. 677). If Rorschach signs such as the $F+$ score can give us information about ego strength, emotional control, and other personality factors psychiatrically said to be important for recovery (9, 69), the Rorschach test warrants the large amount of attention it has received. It is well to remember that the reason for the many claims for the Rorschach may not be solely the sensitivity of the test to important personality variables. Some influence can be attributed to the freedom of interpretation permitted, the frequent immunity of the results to statistical treatment, and the large number of discrete indices whose sheer number is bound to produce statistical significances if the .05 level of confidence is accepted.

Psychoses. The leading claim for the prognostic efficacy of the Rorschach technique with psychotics is that voiced by Piotrowski in a series of eight articles (98, 118, 123, 124, 125, 126, 127, 128). Three descriptions of the crucial personality differences between outcome groups have been propounded: (a) one based on functioning at or below capacity (124), (b) the second based upon a distinction between "emotional" and "intellectual regression" (126), and (c) the third based on a change in either the "variable" or "constant" personality traits (127). Evidence of subnormal functioning at the patient's optimal level of capacity, intellectual regression, and change in the personality traits normally expected to remain constant were held to be unfavorable for improvement. This prognosis was thought to apply regardless of type of therapy (118).

Apparently Piotrowski tried to employ an objective method of revealing those personality characteristics related to outcome. He did this by stipulating retrospectively those quantitative Rorschach signs which showed predictive value. The logic and the substance of this application of experimental procedure are questionable. Aside from such methodological flaws as the omission of Yates's correction when using the chi-square test with small samples (123), there is an almost complete neglect of cross validation of empirical signs. All but one of this group of studies are *ex post facto* in nature, the relationship between predictive test and outcome having been determined after both were known. When this is the case, the experiment must be repeated to determine whether statistically significant results may not be due to sampling error. This need for cross validation becomes clear in view of the considerable fluctuation among the empirical indices reported by Piotrowski in his several studies.

The one attempt at a confirmatory experiment consisted of "blind" prognoses which were reported to be based upon information gained from earlier experiments (126). The Rorschach criteria employed to predict improvement were: "(A) a moderate deviation of personality from the norm of healthy adults, and (B) inhibition of responsiveness to the environment caused by the schizophrenic's fear that the disease process will prevent him from maintaining his former adjustments" (126, p. 808). These criteria, however, do not appear to conform very closely to Piotrowski's previous generalizations nor do they conform to the previous empirical signs (see Table 1). In view of this divergence, it is doubtful whether this study can be considered confirmatory. It is difficult to understand why Piotrowski failed to utilize the previously found empirical indices or how the nebulous criteria he preferred could be objectively handled.

Piotrowski's latest study (128) appears to constitute still another revision of his position. Fifteen signs, most of them new in structure and/or content, were weighted to form a battery which was, again retrospectively, differential in respect to outcome. This new battery remains to be validated.

Numerous authors have expressed support for Piotrowski's descriptions of the personality traits related to prognosis. Their studies, however, have either lacked evidence for their conclusions (11, 82, 117, 168) or employed different, if not opposing, signs from those Piotrowski used (61, 111, 164). Only three of the studies supporting Piotrowski's views (including those done by Piotrowski) stated the statistical significance of the findings, and two of these (111, 123) disregarded Yates's correction for small entries in chi-square tests (see Table 1).

In contrast, there are two studies in which the authors subordinated interpretation to results. Thus, Graham (57) found that *chiaroscuro* responses indicated favorable prognosis but preferred not to identify this index with any hypothesized personality variable, although

Piotrowski (126) felt chiaroscuro was implied in his signs and therefore in his theory. McCall (102) departed even more radically from traditional findings and interpretations. Using rating scales of various personality and perceptual dimensions, he found that the few prognostic indicators which did appear pertained to various aspects of the *content* of Rorschach responses which receives relatively little emphasis in orthodox methods of scoring and interpretation.

In even greater contrast, some investigators have reported an absence of prognostic value for the Rorschach technique (44, 80, 130, 131, 135, 136, 143, 171). The most impressive of these studies (44, 131) consist of thorough attempts to verify the prognostic efficacy of the signs reported by Piotrowski and others. The fact that these rigorous confirmatory studies tend to be negative militates against the acceptance of the rather unconvincing positive conclusions which have been advanced.

Neuroses. Of the five articles found reporting prognostic efficacy of the Rorschach specifically for neurotics, two were no more than hints for future research (46, 85). Another (43) was nullified when cross validation proved the previously reported sign to have no validity.² A report by Dickson (36) lists six signs found to differentiate significantly between outcome groups. Siegel (156), on the other hand, found that a very low percentage of signs showed prognostic value in both an exploratory and cross-validating study. Those signs that did have general predictive power differed from those of Dickson. The relationships between "trait-syndromes judged from the entire protocol" and improvement in psychotherapy were reported to be slight, although statistically significant. None of the blind predictions of improvement was significant.

Roberts (136) has tried to verify previous prognostic claims for the Rorschach in a homogeneous group of patients who seem best to fit the category of neurotics. None of the eleven indices derived from the literature proved prognostic. This might have been expected since several of Roberts' hypotheses originated in studies specifically dealing with psychotics.

Perhaps related to a discussion of neuroses is a study of the treatability of personality difficulties related to sexual promiscuity (18). Rorschach signs relevant to treatability, generally similar to those found by Dickson, were derived from criterion outcome groups, the membership of each group being restricted to women who were judged to show "intrapsychic conflict." These criteria maintained their prognostic power when applied to another sample of promiscuous women.

Problem children. None of the six Rorschach reports on the prognosis of problem children can be considered satisfactory. Four of the articles (90, 152, 153, 154) read more like advertisements than experi-

² SIEGEL, S. M. Personal communication, 1950.

ments. The fifth (155) presents evidence for the alleged prognostic power of twelve Rorschach factors in such form that it cannot be evaluated, and the sixth (141) does not specify the Rorschach signs used.

Alcoholism. There have been two studies concerned with the prognostic use of the Rorschach for alcoholics. These are in agreement that "extraversion" is a favorable sign and "rigidity" unfavorable. The agreement between findings is clear, however, only on the level of interpretation. The article by Sillman (157) provided no evidence for the prognostic claims and did not specify the experimental indices basic to the interpretations. It also appears that the findings were not derived wholly from the Rorschach record. The articles by Billig and Sullivan (12, 13), on the other hand, are replete with data. But there is some inconsistency with respect to the prognostic indices between the two articles (12, p. 126; 13, p. 572), and statistical evaluation of either the individual indices or the total battery is lacking.

Apparently all that can be said is that we have a few hints as to what personality factors may be important for prognosis in alcoholism, the most likely being "extraversion" and "rigidity." In concurrence with these hints, Hoch (73) has reported that in 200 cases of alcoholic psychosis extraverts were three times as frequent as introverts and had a prognosis three times as favorable. In this study no attempt seems to have been made to define and measure extraversion objectively. Other suggestions of prognostic personality factors for alcoholics have been even less objective and either not subjected to confirmatory study (88, 109, 115) or not confirmed when reinvestigated (52, 166).

Somatic illness. The three studies of the prognostic use of the Rorschach for somatic illness are based on the assumption that there is a close relationship between the patient's mental attitude and the ensuing progress of the physical disease. It may or may not be logical to assume that the same mental attitudes lead to improvement in different somatic conditions. In any case, the suggestions provided for each of the different types of illness differ irreconcilably. Thus, Harris and Christiansen (67)³ found among cases showing delay in recovery from physical disease, operation, or accident that poor outcome in psychotherapy was directly related to responses based on the perception of color (color unfavorable), while Ellis and Brown (39) found declining tuberculosis patients had less emotional contact with the environment (low sum C) than did recovering patients (color favorable). Levi (93, 94) claimed that the non-rehabilitated physically handicapped differed from the rehabilitated in showing a high percentage of anatomy responses; but Harris and Christiansen (67) reported anatomy-sex responses favorable for psychotherapeutic recovery.

³ It was felt that the Harris and Christiansen study of psychotherapeutic outcome may be fairly compared with those of the course of physical disease since the modification of physical symptoms was among the criteria of improvement.

On the other hand, there is some consistency among the reports. High *M* scores were found favorable and rejection of the stimulus material unfavorable by both Harris and Christiansen (67) and Ellis and Brown (39). Here also confirmatory studies are desirable.

Mental deficiency. The one Rorschach study of the prognosis of mental defectives presents another striking example of how inadequate the Rorschach test appears to be when subjected to predictive checks. Using employed mental defectives, Sloan (161) tested seven criteria (see Table 1) suggested by Beck as prognostic of good adjustment. Beck's criteria significantly differentiated the two outcome groups, but in the unhypothesized direction. Patients who were successful in staying out on wage placement had a reliably greater number of *disagreements* with Beck's criteria than did those who were returned as failures. Sloan felt that "such an absurdly contradictory finding must be attributed to inadequacy of sufficient differentiation in the two groups and, therefore, should not be accepted at face value" (161, p. 306). It is difficult to understand why Sloan felt justified in discounting this finding while advocating as an important prognostic clue his qualitative index of color and shading shock based on the same inadequately differentiated groups.

General paresis. The Rorschach technique has also been said (32) to be of prognostic value for general paresis. Although prognostic criteria have been outlined, no definitive study has been undertaken to demonstrate this relationship.

Suicide. There have been several Rorschach studies of suicidal tendencies, the most thorough being those of Hertz (71, 72), but these will not be discussed here since they are more often diagnostic than prognostic. In order to orient the study of suicide around prognosis, it will be necessary to obtain more cases such as the one reported by Rabin (129) in which a Rorschach syndrome combining color and shading shock *preceded* murder and attempted suicide. It also seems necessary to differentiate more carefully between serious and non-serious attempts at suicide (40).

Summary. The foregoing review of the prognostic utility of the Rorschach has failed to disclose any very encouraging concordance among studies for any diagnostic category. A considerable number of the positive claims cited in the literature appear to be due to an uncritical attitude concerning validity. Many others may easily be due primarily to chance. When both of these variables are controlled, very few positive results remain.

FREE DRAWING TEST

Fiedler and Siegel (43) have presented evidence that psychoneurotics who draw the face and head of a figure "primitively" have a relatively poor chance for improvement. The authors interpreted poor perform-

ance in drawing the face as "indicative of inability to form that interpersonal relationship between patient and therapist which is the necessary context of the therapeutic process." Before such an interpretation can be accepted, this study should be repeated and more data relating drawing performance to personality assembled.

MINNESOTA MULTIPHASIC PERSONALITY INVENTORY

Although the MMPI is somewhat less subject than is the Rorschach to the fluctuations brought about by an infinity of ready-made indices and interpretations, there is enough freedom to permit considerably more exploration than confirmation and more disagreement than agreement. This fact can be seen from the results summarized in Table 1. It is apparent that there is much conflict among the MMPI indices claimed to be prognostic of favorable outcome in different studies.

Psychoses. High scores on the most frequently cited scale (*Sc*) have been held about equally often to be associated with improvement (20, 60, 66) and with unimprovement (42, 64, 119, 120). Even more striking is the apparent reversal of attitude on the part of Harris regarding the signs for unimprovement. Whereas initially he (64) felt that poor prognosis was associated with a pattern of high scores in the psychotic scales, including psychasthenia, he and others (66) later reported that among chronically ill patients the unimproved averaged within the normal range before treatment on all the MMPI scales, in contrast to the improved who had very high scores either on several of the scales or only on the depression scale. Unfortunately, the data in support of either of these conclusions are not available nor was the significance of either trend reported. Although data on the accuracy of blind prognostic sorting have been presented (66), the criteria on which it was based are unclear, a situation parallel to Piotrowski's blind sorting (126).

In light of evidence from other studies (175, 183) on the importance of chronicity in prognosis and of additional information concerning the different types of patients employed in the studies by Harris and Feldman,⁴ it was thought that much of the apparent disagreement among the MMPI reports might be attributed to radical differences among the patients in chronicity. This interpretation seems plausible, since at least two of the three studies that reported high scores as generally favorable employed patients of long duration of illness, while at least two of the four studies reporting high scores as generally unfavorable dealt with patients of short duration of illness and/or eliminated patients showing false normal scores. There remains, however, some question of the validity of this explanation of divergent findings. Unfortunately, many of the studies have failed to provide sufficient information in regard to both the conditions of the experiment and statistical evaluation of the

⁴ HARRIS, R. E. Personal communication, 1951.

results. Further, some findings seem to be inconsistent even when chronicity is considered. Carp (20) studied a group of patients of apparently wide range of chronicity (see Table 1). One would expect this group to be differentiated in a manner similar to acutely ill patients (high scores unfavorable), since most of the recovered cases should have been acutely ill. Harris' (66) sorting based on criteria "developed in other studies" (presumably studies dealing with acutely ill patients) would not be expected to be so accurate with chronically ill patients if the hypothesized reversal in prognostic criteria with difference in chronicity applies (see Table 1).

It is highly questionable whether one is justified in concluding that it has been demonstrated that the scores of any of the regular MMPI scales can prognosticate the outcome of therapy for psychotics (65). Probably any such generalization must take into account the role of chronicity (among other variables).

Among the most thorough investigations of the prognostic use of the MMPI are those of Feldman (41, 42) and Pearson (119, 120); both series of studies involved cross validation (see Table 1). Feldman empirically developed a new MMPI scale to be used for prognosis independent of diagnosis. This scale has been applied to more subjects than are customarily used in prognostic experiments and has proven effective, although some of the efficiency may be due to the combination of unlike diagnostic categories. Further work in prognosis should employ this scale or, just as desirable, previously gathered data should be re-analyzed with it. Pearson was concerned only with immediate response of schizophrenics to ECT, regardless of final outcome. Subsequent study of the prognostic efficacy of the MMPI signs for later outcome should help to establish the relationship between different criteria of outcome.

Neuroses. Two scales have been applied to neurotics to determine prognostic utility (42, 148). Only the abbreviated *Ps* scale of Feldman (42) seems effective, but this scale lacks cross validation.

Somatic illness. A number of investigators (67, 137, 138, 139) have reported that cases of delayed recovery from physical disease, operation, or accident are characterized by abnormally high MMPI scores, especially on the neurotic scales. The clinical impression of these patients seems to be that they are neurotic. Ruesch *et al.* (139) further described the delayed recovery group as containing the hysterical or anxiety type (women) and the dependent type (men).

Harris and Christiansen (67), going one step further, found high MMPI scores prognostically unfavorable in psychotherapeutically treated cases of delayed recovery. In order to identify those attitudes contributing to high scores, items were grouped into 35 subscales, eight of which showed significant differences between prognostic groups. The attitudes represented by these eight differentiating subscales were thought to be paranoid and psychopathic, indicating that there may

be subclinical psychotic or psychopathic trends in the delayed recovery group with poor prognosis. These findings at present lack cross validation and may have been superseded in importance by Feldman's (42) use of the *Ps* scale with the same group of subjects (see above).

Summary. Prognostic studies of psychotics using the MMPI exhibit a large amount of disagreement among conclusions. Evidence indicates that some of this disagreement can be attributed to differences in the type of patients studied. Investigation directly on this question seems essential to establishing a consistent and comprehensive picture of prognosis.

The prognostic MMPI studies of somatic illness seem essentially in agreement and promise considerably greater utility than do the Rorschach studies in this field.

THEMATIC APPERCEPTION TEST

So far as the reviewer has been able to discover, the TAT has not been employed to determine prognostic criteria of remission, even though there has been an attempt to demonstrate with case studies how the TAT could be of prognostic value (81). Hartman (68) determined biserial correlations between the psychiatric rating of good behavior prognosis and 56 TAT categories for 35 delinquent boys. He found eight "significant" ($p < .06$) correlations. However, a third of these could be expected by chance, and correlations between ratings of prognosis from blind analysis of the TAT and either the experimental or psychiatric rating were .15 or lower.

Masserman and Balken (108) reported an "occasional significance" of TAT phantasies for prognosis. They felt that the type of change in phantasy with psychiatric interviews could show the abatement of conflicts and the development of insight. Conversely, poor prognosis would result from untouched intense intrapsychic conflicts or unconscious resistance as expressed in recurring phantasies about sick people who cannot be cured. All in all, it does not appear that objective criteria have been found through which the TAT can be of prognostic use, even though indications that it may be useful are available.

MOSAIC TEST

The Mosaic Test has also been asserted to be of prognostic utility, but again there are no studies that have demonstrated its value in this area. Diamond and Schmale (35) have described a technique that is an offshoot of the Mosaic Test to detect the "psychological color blindness" of the schizophrenic. They felt this approach to the thinking disorder of schizophrenics would lead to more accurate prognostic stand-

ards than those afforded by sorting tests, but as yet this relationship remains undemonstrated.

MEASURES OF AGGRESSION

Recently two studies have reported prognostic value for measures of aggression. Extrapunitiveness and intropunitiveness measured in terms of acts of aggression or "accident" directed toward others or toward oneself were found to be significantly related to outcome for psychiatric patients and for schizophrenics alone (3). Extrapunitive behavior was unfavorable, intropunitive behavior favorable. Scores of the direction of aggression derived from the Picture Frustration Study were not found to be prognostic, although it was suggested that an antithetical pattern between the two measures (overt and P-F) may point to a poor prognosis, especially when the overt measure is extrapunitive (5). These clues are interesting but require verification and more refined methods of measurement than those employed.

TESTS OF MENTAL ABILITY

Psychoses. Attempts to determine the role of intellectual capacity in clinical prognosis have been more objective than investigations of other phases of personality, mainly because the accepted tests of intelligence are relatively standardized in scoring and interpretation. On the other hand, the extent to which different tests measure common aspects of ability is unclear. There is also a large amount of disagreement in the general findings reported. Often it is claimed that psychotics manifesting the greater intellectual power are more likely to improve (16, 21, 57, 89, 92, 98, 105, 143, 144, 145, 174, 175, 181), but just as frequently no relationship is found (45, 57, 66, 121, 130, 143, 144, 165, 181, 183) or else those who perform less efficiently on ability tests show the better outcome (91, 92, 106, 107, 116, 133, 145, 175, 182, 183).

Evaluation of these reports must take into consideration such procedural flaws as the lack of statistically significant evidence (91, 106, 107, 116, 128, 133, 145). Another factor to be considered is the particular tests employed, or the supposedly different abilities involved. Unfortunately, it is hard to deal with this factor in view of our inadequate knowledge of the structure of general or specific abilities in the mentally ill. This inadequacy is not remedied by the various interpretations accompanying prognostic claims.

If it is assumed that the abilities measured in different studies, despite differences in instruments and techniques, are similar, the contradictions in results are even more disturbing than those found for less objective tests. When one uses the IQ as a prognostic indicator, the index is fairly well defined, and ambiguity in the results cannot easily be attributed to vagaries in interpretation.

Although there is evidence that some of the obtained differences in

the prognostic value of intelligence tests may relate to diagnostic differences (106, 107, 174), to the use of different types of therapy (144, 145, 181), and to different follow-up periods (145, 175, 183), these factors are not sufficient to eliminate all of the existing inconsistencies. Probably the most hopeful attack on this problem has been the attempt to fractionate groups of patients into like-minded or like-structured subgroups, each of which is characterized by different prognostic indices. Prognostic studies carried out in the first Columbia Greystone Project indicated an association of *poor* test performance with favorable outcome (91, 182, 183). It was thought that the chronic status of these patients might be a crucial factor in explaining the inverse prognostic relationship. Consequently, the prognostic implications of the Complex Reaction Time Test for the chronically ill (long duration) Columbia Greystone patients were compared with those for acutely ill (short duration) psychotics (175). The fact that chronic psychotics who performed poorly and acutely ill psychotics who performed well later improved offers a possible basis for resolving the apparent contradictions in the literature. This concept would also help to explain the differing prognoses for metrazol and insulin shock therapies (145, 181), since the more chronically ill are usually given metrazol. It may be that temporal duration of illness is not a very satisfactory measure of chronicity, since it is difficult to define or measure and since the disease process probably progresses at different rates in different patients. This measure has yielded some degree of success, however, and is a lead worth following up. It should be noted that the above study has not been cross validated, so that the role of chronicity is not definitely established.

Neuroses. Three studies (86, 104, 110) have agreed that high intelligence leads to a favorable outcome in neuroses, but despite this agreement the evidence from these studies is not strong enough to justify the conclusion that a relationship has been demonstrated. Klugman (86) provided no statistical measure of significance and Malamud and Gottlieb (104) reported their findings significant to a lower degree than is usually accepted. In addition, Dickson (36) reports the IQ to be of no prognostic value for neurotics.

Somatic illness. The two studies relating intelligence to outcome of somatic illness differ concerning the prognostic value of the IQ. Harris and Christiansen (67) found no relationship, while Ruesch *et al.* (139) reported that high intelligence is significantly favorable. We lack sufficient evidence to determine whether measures of intelligence can be used prognostically in this connection.

PATTERNS OF ABILITY

The fact that interrelationships among particular abilities or traits may be correlated with tendencies toward adjustment is the rationale for the use of pattern analyses. This approach has been used extensively with the Rorschach technique, greatly increasing the number of appar-

ent differences between groups in exploratory studies. The use of pattern analysis with ability tests provides sorely needed information concerning the relationships of abilities without being dangerously subject to chance "findings" since relatively few combinations exist.

The study of patterns of abilities, although not widely used, has focused largely on patients whose primary difficulty is the loss or lack of ability—the deteriorated and the mentally defective. Two studies (38, 62) of mental defectives have indicated that those whose "performance" scores exceed or equal their "verbal" scores have better prognoses of social adaptability than do those whose verbal scores exceed the performance scores. The same prognostic relationship has been claimed for schizophrenics (8, 98).

Other patterns of abilities investigated for prognostic value include relationships between timed and untimed performance tests (66), delayed and immediate memory (143, 144,) and Stanford-Binet subtests (scatter) (57, 105). In no case is there confirmation of the various claims.

MISCELLANEOUS TECHNIQUES

There have been many other suggestions of psychological criteria for prognosis, most of which do not fall conveniently into the previously mentioned categories of tests. Many of these criteria are not objective, but nonetheless merit consideration.

Recently Voth followed up a suggestion by Sexton (150) that visual autokinesis may be of prognostic value. Voth (169) found that intermediate degrees of apparent movement were favorable for psychotics. Although more subjects were employed in this study than is typical in this field, the results require validation.

A number of reports (4, 6, 7, 8, 52, 54, 58, 73, 149, 159, 166, 172) have related outcome to clinical estimates of personality or behavioral traits. By and large such approaches leave much to be desired in the way of controls and reproducibility. It also appears that among comparable studies there are definite disagreements. Thus Gray (58) and Wender (172) differ as to whether patients showing maximal participation in group psychotherapy will be most or least likely to improve. Tillotson and Fleming (52) indicated that empirical findings may be less reliable than they appear at first glance. After an attempt to confirm the prognostic efficacy of a number of personality traits they had found retrospectively important for chronic alcoholism (166), they concluded that "the outcome of treatment of chronic alcoholism has little apparent relation to sociological or personality traits" (52, p. 744).

One of the most "experimental" investigations of this sort is the study of Peters (121) in which rating sheets filled out from interviewers'

records permitted a frequency count of trait names. Twelve traits were found to be predictive of outcome. Peters interpreted these as showing that a "generally heightened activity level" and a state of "integrity and dominance of the cortical centers" were favorable to improvement. Whether or not this conclusion turns out to be valid, the approach is an admirable attempt to introduce some quantification into what is too often but a potpourri of impressions.

A number of prognostic reports have arisen from studies of the results of nondirective therapy. Snyder (162) with only five cases, found the one unsuccessful case expressed fewer feelings and also more negative and ambivalent attitudes toward the counselor than did the four successful cases. Blau (15) developed a different attitudinal index, one based on statements referring to the self during the first interview. He found that the greater the number of positively-valenced and ambivalently-valenced self statements, and the fewer the negatively-valenced self statements, the better the prediction for therapeutic success.

There have also been some reports based upon psychoanalytic theory. Knight (88) stated that those alcoholic individuals well developed in the "second anal state" offered the best material for therapy. Piers (122) claimed that only the "oral types" of schizophrenic responded well to insulin treatment, "anal types" responding poorly. It was tentatively proposed that the decisive psychodynamic factor in insulin treatment was the guiltless fulfillment of an immense oral craving. Obviously, these reports should be considered merely suggestive, since they remain unsupported by evidence or operational definitions.

Skottowe (160) has pointed out three types of schizophrenics that he felt called for differently arranged and proportioned forms of therapy. The dys-symbolic type is unable to formulate "conceptual thoughts" upon personal topics or to discriminate linguistically the gradations of his emotions, even though he is in a state of clear consciousness and can use words for perceptual thinking. This type of case, according to Skottowe, fails to recover with shock therapy. The dyskinetic—those with "disorders of motility"—recover well, while the simple paranoid type becomes accessible to further needed psychotherapy through the use of shock therapy. Thomas (167) provided data from 32 dys-symbolic cases to support Skottowe's thesis; not one of these cases recovered. A better test of this hypothesis is needed, as well as an enumeration of objective criteria of classification of these three types.

Perhaps the most thorough attempt at determining prognosis has been the work of Wittman with the Elgin scale. This scale contains 30 weighted factors thought to be prognostic for functional psychoses

(176). These factors involved numerous psychological and clinical characteristics derived from studies previous to 1941. The scores of schizophrenics were found to be bimodally distributed on this scale, those of manic-depressives skewed (179). In the schizophrenic group there was a correlation of prognostic score with diagnostic subcategory: (a) the hebephrenic and simple types had poor prognosis; (b) the catatonic and undetermined groups were bimodally distributed with regard to prognosis; and (c) the paranoid subtype did not fit well into the scale. Prognosis made by the Elgin scale was considerably better than that made by the clinical staff. However, the probability of this advantage occurring by chance was not indicated, the comparison being presented in percentages. Wittman suggested on the basis of her analysis that the duration of psychosis as a criterion for predicting improvement is an artifact rather than a true criterion. Patients with a long duration of illness show poorer prognosis because they are basically ill, not the reverse.

Wittman (177) has more recently found a significantly larger number of cases of "constitutional" schizophrenics (those who had been maladjusted as children) than "functional" showed a high positive (unfavorable prognostically) score on the Elgin rating scale, high weighting in heboid regressive features, poor institutional development, and lack of improvement with shock therapy. She thinks that longitudinal studies are necessary for prognosis, and accepts Langfeldt's (92) distinction between process schizophrenia and schizophreniform types of psychosis. Data have been reported by Sarbin which further emphasize the usefulness of this scale (142). In his study the weights for the scale items were derived empirically, rather than arbitrarily assigned, and similar results were obtained.

There have been quite a few indications of the value of morphological indices for prognosis. Such factors were introduced by Kretschmer's report that dementia praecox patients were most often of asthenic physique and manic-depressives of pyknic physique. This led to the correlative hypothesis that in the functional psychoses pyknics have a better prognosis than do asthenics. Studies that report morphological differences in outcome for mental disorders have usually stressed the better prognosis for patients of pyknic than for those of asthenic habitus (22, 25, 53, 56, 79, 83, 84, 95, 96, 104, 112, 163), but some investigators have failed to find any relationship between outcome and body type (98, 104, 133, 183) or have found the "pyknotic" unfavorable (105).

DISCUSSION

In order for the investigation of prognostic indices to become reasonably effective, studies in this field will have to meet at least four criteria:

1. The conditions of the experiment must be specified. This means that the patient population must be described in all pertinent details, the conditions of therapy must be stipulated, and objective criteria of outcome must be presented. Many of the studies in the literature have failed to specify these crucial characteristics.

2. The experiment must deal with relatively homogeneous populations and conditions. Numerous studies (106, 107, 144, 145, 174, 175, 181, 183) have demonstrated that the prognostic indices applying for groups differing in diagnosis, chronicity, or therapy may differ. The investigation of broad categories such as "psychiatric patients" may be said to be just as meaningful as studies of very delimited groups. But since the differential prognoses of manic-depressives and schizophrenics, for example, are already known, it would appear more useful to discover the prognostic indices for each subgroup. If investigations were carried out under homogeneous conditions we could eventually build up the composite picture, an accomplishment much less likely to result from studies in which these variables are uncontrolled.

3. Findings should be reported in terms amenable to statistical evaluation. Even though most investigators have presented evidence in quantitative form, all too frequently they have not made tests of statistical significance. Such information in terms of probabilities is necessary to establish the confidence that can be put in the findings. Furthermore, tests of significance have often been misinterpreted. Cochran and Cox (26) clearly state the misconceptions which can arise from a misapplication of the laws of probability:

In order that the *F*- and *t*-tests be valid, the tests to be made in an experiment should be chosen before the results have been inspected. The reason for this is not hard to see. If tests are selected *after* inspection of the data, there is a natural tendency to select comparisons that appear to give large differences. Now large apparent differences may arise either because there are large real effects, or because of a fortuitous combination of the experimental errors. Consequently, in so far as differences are selected just because they seem to be large, it is likely that an undue proportion of the cases selected will be those where the errors have combined to make the differences large. The extreme case most commonly cited is that of the experimenter who always tests, by an ordinary *t*-test, the difference between the highest and lowest treatment means. If the number of treatments is large, this difference will be substantial even when the treatments produce no real differences in effect. It may be shown that with 3 treatments the observed value of *t* will exceed the 5% level in the table about 13% of the time. With 6 treatments the figure is 40%, with 10 treatments 60%, and with 20 treatments 90%. When the experimenter thinks that he is making a *t*-test at the 5% level, he is actually testing at the 13% level, or the 40% level, and so on (26, pp. 67-68).

4. Findings must be subjected to cross validation. Another indication of the lack of statistical sophistication in this field is the large proportion of exploratory findings in relation to the small proportion of confirmatory findings reported in the prognostic literature. Obviously,

exploratory or previously unhypothesized findings must be verified before they can be regarded with confidence.

When these criteria have been met, it will be possible to determine with a much greater degree of accuracy whether psychological tests can have prognostic value and what range of applicability any particular indices may have. Since the majority of prognostic studies have been conducted and interpreted in broad diagnostic, therapeutic, and outcome terms, it has been necessary to evaluate them on this basis. It seems very likely, however, that the broad categories employed in this review will not prove feasible in future prognostic work.

To obtain the prognostically homogenous conditions that seem basic to any future research, there may be a need (possibly common to all fields of psychology) for an over-all coordinating and evaluating organization. Cattell (24) has pointed out the need for a quorum vote of policy regarding the factors of mental organization. Others (50, 76) have proposed a general psychological research exchange. Similarly, in the field of prognosis a coordinating committee could be of service in organizing more meaningful experiments and encouraging reporting of results in quantitative form.

Thorough discussions of the logical problems involved in prediction and critical considerations of basic research procedures are available (30, 31, 37, 59, 74, 75, 103, 114, 141, 142). It would be highly desirable if research workers were more familiar with works of this sort and were guided to a greater extent by the organization they provide.

SUMMARY

There is no dearth of prognostic studies attempting to relate psychological test performance to eventual outcome in psychopathy. However, many of the individual studies fail to demonstrate empirical justification for the conclusions drawn. Further, comparisons among studies reveal little agreement among findings. With the possible exception of certain findings that may prove to be valid, it appears that previous research on prognosis can tell us only that it is not possible to discover reliable prognostic criteria without controlling variables to a much greater extent than has been customary. When investigators in this field recognize the need for certain experimental and evaluative procedures we may hope to discover what, if any, prognostic power psychological tests may have.

BIBLIOGRAPHY

1. ABEL, T. M. The Rorschach Test and school success among mental defectives. *Rorschach Res. Exch.*, 1945, 9, 105-110.
2. ADLER, D. L. Normal vs. schizophrenic perception of similarities. *Psychol. Bull.*, 1942, 39, 507-508. (Abstract)

3. ALBEE, G. W. Patterns of aggression in psychopathology. *J. consult. Psychol.*, 1950, **14**, 465-468.
4. ALBEE, G. W. The prognostic importance of delusions in schizophrenia. *J. abnorm. soc. Psychol.*, 1951, **46**, 208-212.
5. ALBEE, G. W., & GOLDMAN, R. The Picture Frustration Study as a predictor of overt aggression. *J. proj. Tech.*, 1950, **14**, 303-308.
6. ALEXANDER, F., & FRENCH, T. M. *Psychoanalytic therapy*. New York: Ronald, 1946.
7. ARNOT, R., TALBOT, B., & GREENBLATT, M. One to four year follow-up of 205 cases of bilateral prefrontal lobotomy. In M. Greenblatt, R. Arnot, & H. Solomon (Eds.), *Studies in lobotomy*. New York: Grune & Stratton, 1950.
8. BALINSKY, B. Factors in the vocational adjustment of schizophrenics after mental hospital discharge. *J. clin. Psychol.*, 1947, **3**, 341-349.
9. BECK, S. J. Rorschach F plus and the ego in treatment. *Amer. J. Orthopsychiat.*, 1948, **18**, 395-401.
10. BELLAK, L. *Dementia praecox*. New York: Grune & Stratton, 1948.
11. BENJAMIN, J. D. A method for distinguishing and evaluating formal thinking disorders in schizophrenia. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia*. Berkeley: Univer. of California Press, 1946. Pp. 65-90.
12. BILLIG, O., & SULLIVAN, D. J. Prognostic data in chronic alcoholism. *Rorschach Res. Exch.*, 1942, **6**, 117-127.
13. BILLIG, O., & SULLIVAN, D. J. Personality structure and prognosis of alcohol addiction. *Quart. J. Stud. Alc.*, 1943, **3**, 554-573.
14. BLAIR, D. Prognosis in schizophrenia. *J. ment. Sci.*, 1940, **86**, 378-477.
15. BLAU, T. H. Report on a method of predicting success in psychotherapy. *J. clin. Psychol.*, 1950, **6**, 403-406.
16. BOLLES, M. M., ROSEN, G. P., & LANDIS, C. Psychological performance tests as prognostic agents for the efficacy of insulin therapy in schizophrenia. *Psychiat. Quart.*, 1938, **12**, 733-737.
17. BOSS, M. Psychopathologie des Traumes bei schizophrenen und organischen Psychosen. *Z. ges. Neurol. Psychiat.*, 1938, **162**, 459-494.
18. BRADWAY, K. P., LION, E. G., & CORRIGAN, H. G. The use of the Rorschach in a psychiatric study of promiscuous girls. *Rorschach Res. Exch.*, 1946, **10**, 105-110.
19. BROWN, M. A. Alcoholic profiles on the Minnesota Multiphasic. *J. clin. Psychol.*, 1950, **6**, 266-269.
20. CARP, A. MMPI performance and insulin shock therapy. *J. abnorm. soc. Psychol.*, 1950, **45**, 721-726.
21. CARP, A. Performance on the Wechsler-Bellevue Scale and insulin shock therapy. *J. abnorm. soc. Psychol.*, 1950, **45**, 127-136.
22. CARTER, A. B. The prognostic factors of adolescent psychoses. *J. ment. Sci.*, 1942, **88**, 31-81.
23. CATTELL, R. B. Temperament tests in clinical practice. *Brit. J. med. Psychol.*, 1937, **16**, 43-61.
24. CATTELL, R. B. The measurement of adult intelligence. *Psychol. Bull.*, 1943, **40**, 153-193.
25. CHASE, L. S., & SILVERMAN, S. Prognostic criteria in schizophrenia; a critical survey of the literature. *Amer. J. Psychiat.*, 1941, **98**, 360-368.
26. COCHRAN, W. G., & COX, G. M. *Experimental designs*. New York: Wiley, 1950.
27. COHEN, L. H. The therapeutic significance of fear in the metrazol treatment of schizophrenia. *Amer. J. Psychiat.*, 1939, **95**, 1349-1357.
28. COOK, L. C. Has fear any therapeutic significance in convulsion therapy? *J. ment. Sci.*, 1940, **86**, 484-490.
29. COWEN, E. L., & COMBS, A. W. Fol-

- low-up study of 32 cases treated by nondirective psychotherapy. *J. abnorm. soc. Psychol.*, 1950, 45, 232-258.
30. CRONBACH, L. J. Statistical methods applied to Rorschach scores: a review. *Psychol. Bull.*, 1949, 46, 393-429.
31. DAILEY, C. A. The clinician and his predictions. *J. clin. Psychol.*, 1951, 7, 270-273.
32. DAVIDSON, G. M., & CONKEY, R. C. The Rorschach Test and the questions of "prognosis" and "recovery" in syphilitic meningo-encephalitis. *Psychiat. Quart.*, 1950, 24, 243-258.
33. DEARBORN, G. V. N. The determination of intellectual regression and progression. *Amer. J. Psychiat.*, 1926-7, 6, 725-741.
34. DENKER, P. G. Results of treatment of psychoneuroses by the general practitioner: a follow-up study of 500 patients. *Arch. Neurol. Psychiat.*, Chicago, 1947, 57, 504-505.
35. DIAMOND, B. L., & SCEMALE, H. T. The Mosaic Test. 1. An evaluation of its clinical application. *Amer. J. Orthopsychiat.*, 1944, 14, 237-250.
36. DICKSON, L. P. The Rorschach Test as an indicator of prognosis in psychoneurosis. Unpublished master's thesis, Catholic Univer., 1949.
37. DUMAS, F. M. The objective evaluation of prognostic proficiency. *J. gen. Psychol.*, 1950, 43, 131-135.
38. EARL, C. J. C. A psychograph for morons. *J. abnorm. soc. Psychol.*, 1940, 35, 428-448.
39. ELLIS, R. W., & BROWN, G. G. The nature of Rorschach responses from pulmonary tuberculosis patients. *J. clin. Psychol.*, 1950, 6, 298-300.
40. FARBEROW, N. L. Personality patterns of suicidal mental hospital patients. *Genet. Psychol. Monogr.*, 1950, 42, 3-79.
41. FELDMAN, M. J. A prognosis scale for shock therapy. *Amer. Psychologist*, 1948, 3, 348. (Abstract)
42. FELDMAN, M. J. Some scales to predict and evaluate the results of several types of shock therapy. Unpublished doctoral dissertation, Univer. of California, 1949.
43. FIEDLER, F. E., & SIEGEL, S. M. The Free Drawing Test as a predictor of non-improvement in psychotherapy. *J. clin. Psychol.*, 1949, 5, 386-389.
44. FILMER-BENNETT, G. T. A study of prognostic indices in the Rorschach records of hospitalized patients. Unpublished doctoral dissertation, Univer. of Pittsburgh, 1950.
45. FISHER, K. A. Changes in test performance of ambulatory depressed patients undergoing electro-shock therapy. *J. gen. Psychol.*, 1949, 41, 195-232.
46. FISHER, S. Rorschach patterns in conversion hysteria. *J. proj. Tech.*, 1951, 15, 98-108.
47. FISHER, S. The value of the Rorschach for detecting suicidal trends. *J. proj. Tech.*, 1951, 15, 250-254.
48. FISHER, S., & FISHER, RHODA. Value of isolation rigidity in maintaining integration in seriously disturbed personalities. *J. Pers.*, 1950, 19, 41-47.
49. FISHMAN, S. An investigation of the relationship between references to the self-concept of above the knee amputees, and their adjustment to leg prostheses. *Amer. Psychologist*, 1949, 4, 256. (Abstract)
50. FISKE, D. W. Must psychologists be experimental isolationists? *Amer. Psychologist*, 1947, 2, 23, 28.
51. FLANAGAN, J. C. (Ed.) The Aviation Psychology Program in the Army Air Forces. *AAF Aviat. Psychol. Res. Rep.* No. 1. Washington: U. S. Government Printing Office, 1948.
52. FLEMING, R., & TILLOTSON, K. Further studies on the personality and sociological factors in the prognosis and treatment of chronic alcohol-

- ism. *New Engl. J. Med.*, 1939, 221, 741-745.
53. GILDEA, E. F., & MAN, E. B. Methods of estimating capacity for recovery in patients with manic-depressive and schizophrenic psychoses. *Amer. J. Psychiat.*, 1943, 99, 496-506.
54. GLATZER, H. T. Selection of mothers for group therapy. *Amer. J. Orthopsychiat.*, 1947, 17, 477-483.
55. GOLDSTEIN, K. The significance of special mental tests for diagnosis and prognosis in schizophrenia. *Amer. J. Psychiat.*, 1939, 96, 575-588.
56. GOTTLIEB, B. S. Prognostic criteria in hebephrenia; the importance of age, sex, constitution and marital status. *Amer. J. Psychiat.*, 1940, 97, 332-341.
57. GRAHAM, V. T. Psychological studies of hypoglycemia therapy. *J. Psychol.*, 1940, 10, 327-358.
58. GRAY, W. Group psychotherapy in a state hospital. *J. nerv. ment. Dis.*, 1948, 108, 485-495.
59. GUILFORD, J. P., & MICHAEL, W. B. *The prediction of categories from measurements: with application of personnel selection and clinical prognosis.* Ann Arbor: Edwards Bros., 1949.
60. HALES, W. M., & SIMON, W. Minnesota Multiphasic Personality Inventory patterns before and after insulin shock therapy. *Amer. J. Psychiat.*, 1948, 105, 254-258.
61. HALPERN, F. Rorschach interpretation of the personality structure of schizophrenics who benefit from insulin therapy. *Psychiat. Quart.*, 1940, 14, 826-833.
62. HAMLIN, R. Test pattern of high grade mentally defective girls. *Proc. Amer. Ass. ment. Def.*, 1938, 43, 161-165.
63. HARRIS, I. D., RAPOPORT, L., & RYNERSON, M. A. Observations on asthmatic children. *Amer. J. Orthopsychiat.*, 1950, 20, 490-505.
64. HARRIS, R. E. Measured personality characteristics of convulsive therapy patients: a study of diagnostic and prognostic criteria. *Psychol. Bull.*, 1945, 42, 535. (Abstract) *impr.*
65. HARRIS, R. E. Psychodiagnostic testing in psychiatry and psychosomatic medicine. In M. R. Harrower (Ed.), *Recent advances in diagnostic psychological testing.* Springfield: Thomas, 1950. Pp. 13-29.
66. HARRIS, R. E., BOWMAN, K. M., & SIMON, A. Studies in electronarcosis therapy. III. Psychological test findings. *J. nerv. ment. Dis.*, 1948, 107, 371-376. *impr.*
67. HARRIS, R. E., & CHRISTIANSEN, C. Prediction of response to brief psychotherapy. *J. Psychol.*, 1946, 21, 269-284. *impr.*
68. HARTMAN, A. A. An experimental examination of the Thematic Apperception Technique in clinical diagnosis. *Psychol. Monogr.*, 1949, 63, No. 8 (Whole No. 303).
69. HERTZ, M. R. The role of the Rorschach method in planning for treatment. *Rorschach Res. Exch.*, 1945, 9, 134-146.
70. HERTZ, M. R. The Rorschach method and its significance in the Mental Hygiene Program. In P. L. Harriman (Ed.), *Twentieth century psychology.* New York: Philosophical Library, 1946. Pp. 652-684.
71. HERTZ, M. R. Suicidal configurations in Rorschach records. *Rorschach Res. Exch.*, 1948, 12, 3-58.
72. HERTZ, M. R. Further study of "suicidal" configurations in Rorschach records. *Rorschach Res. Exch.*, 1949, 13, 44-73.
73. HOCH, P. H. Personality factors in alcoholic psychosis. *Psychiat. Quart.*, 1940, 14, 338-346.
74. HORST, P. *The prediction of personal adjustment.* New York: Soc. Sci. Res. Coun., Bull. 48, 1941.
75. HOTELLING, H., SOROTKIN, P., GUTT-

- MAN, L., & BURGESS, E. W. The prediction of personal adjustment: A symposium. *Amer. J. Sociol.*, 1942, 48, 61-86.
76. IRWIN, F. W. A proposal toward a psychological research exchange. *Amer. Psychologist*, 1948, 3, 91-94.
77. KALINOWSKY, L. B., & HOCH, P. H. *Shock treatments and other somatic procedures in psychiatry*. New York: Grune & Stratton, 1946.
78. KAMMAN, G. R. The Rorschach method as a therapeutic agent. *Amer. J. Orthopsychiat.*, 1944, 14, 21-27.
79. KANT, O. The evaluation of prognostic criteria in schizophrenia. *J. nerv. ment. Dis.*, 1944, 100, 598-605.
80. KATZ, H. Untersuchungen an insulin-behandelten Schizophrenen mit dem Rorschach'schen Formdeutversuch. *Mshr. Psychiat. Neurol.*, 1941, 104, 15-33.
81. KENDIG, ISABELLE V. Projective techniques as a psychological tool in diagnosis. *J. clin. Psychopath.*, 1944, 6, 101-110.
82. KISKER, G. W. A projective approach to personality patterns during insulin-shock and metrazol-convulsive therapy. *J. abnorm. soc. Psychol.*, 1942, 37, 120-124.
83. KLINE, N. S., & TENNEY, A. M. Constitutional factors in the prognosis of schizophrenia. *Amer. J. Psychiat.*, 1950, 107, 434-441.
84. KLINE, N. S., & TENNEY, A. M. Prognosis in topectomies and lobotomies relative to body type. *Arch. Neurol. Psychiat., Chicago*, 1951, 65, 323-325.
85. KLOPPER, W. G. The efficacy of group therapy as indicated by group Rorschach records. *Rorschach Res. Exch.*, 1945, 9, 207-209.
86. KLUGMAN, S. F. Test scores of psychoneurotic returnees before and after convalescent furlough. *J. clin. Psychol.*, 1948, 4, 195-200.
87. KLÜVER, H. Psychology at the beginning of World War II: Meditations on the impending dismemberment of psychology written in 1942. *J. Psychol.*, 1949, 28, 383-410.
88. KNIGHT, R. P. The dynamics and treatment of chronic alcohol addiction. *Bull. Menninger Clin.*, 1937, 1, 233-250.
89. KRIEGMAN, G., & HILGARD, J. R. Intelligence level and psychotherapy with problem children. *Amer. J. Orthopsychiat.*, 1944, 14, 251-265.
90. KRUGMAN, M. Rorschach examination in a child guidance clinic. *Amer. J. Orthopsychiat.*, 1941, 11, 503-511.
91. LANDIS, C., & ERLICK, D. An analysis of the Porteus Maze-Test as affected by psychosurgery. *Amer. J. Psychol.*, 1950, 63, 557-566.
92. LANGFELDT, G. *The prognosis in schizophrenia and the factors influencing the course of the disease*. London: Oxford Univer. Press, 1937.
93. LEVI, J. Rorschach patterns as a tool in predicting success or failure in the rehabilitation of the physically handicapped. *Amer. Psychologist*, 1950, 5, 320-321. (Abstract)
94. LEVI, J. Rorschach patterns predicting success or failure in the rehabilitation of the physically handicapped. *J. abnorm. soc. Psychol.*, 1951, 46, 240-244.
95. LEWIS, A. J. Prognosis in schizophrenia. *Lancet*, 1935, 228, 339-341.
96. LEWIS, N. D. C. What do we know about dementia praecox? *Mem. Hyg.*, 1950, 34, 569-581.
97. LORR, M., WITTMAN, PHYLLIS, & SCHANBERGER, W. An analysis of the Elgin Prognostic Scale. *J. clin. Psychol.*, 1951, 7, 260-263.
98. LOURIE, R. S., PACELLA, B. L., & PIOTROWSKI, Z. A. Studies on the prognosis in schizophrenic-like psychoses in children. *Amer. J. Psychiat.*, 1943, 99, 542-551.

99. LOW, A. A., SINGER, H. D., & MCCORRY, C. L. Life situation tests as aids in psychiatric prognosis. *Amer. J. Psychiat.*, 1939, **96**, 147-162.
100. LOW, A. A., SONENTHAL, I. R., BLAUROCK, M. F., KAPLAN, M., & SHERMAN, I. Metrazol shock treatment of the "functional" psychoses. *Arch. Neurol. Psychiat.*, Chicago, 1938, **39**, 717-736.
101. LUBORSKY, L. B. Psychometric changes during electric shock treatment. *J. nerv. ment. Dis.*, 1948, **107**, 531-536.
102. MCCALL, R. J. Psychometric evaluation of Rorschach records in brain-operated patients. Unpublished doctoral dissertation, Columbia Univer., 1951.
103. MALAMUD, D. I. Objective measurement of clinical status in psychopathological research. *Psychol. Bull.*, 1946, **43**, 240-258.
104. MALAMUD, W., & GOTTLIEB, J. S. Course and prognosis in the psychoneuroses. *J. Mt. Sinai Hosp., New York*, 1942, **9**, 630-639.
105. MALAMUD, W., & RENDER, N. Course and prognosis in schizophrenia. *Amer. J. Psychiat.*, 1939, **95**, 1039-1057.
106. MALAMUD, W., SANDS, S. L., & MALAMUD, I. The involuntional psychoses: a socio-psychiatric study. *Psychosom. Med.*, 1941, **3**, 410-426.
107. MALAMUD, W., SANDS, S. L., MALAMUD, I. T., & POWERS, P. J. P. The involuntional psychoses: a socio-psychiatric follow-up study. *Amer. J. Psychiat.*, 1949, **105**, 567-572.
108. MASSERMAN, J. H., & BALKEN, E. R. The clinical application of phantasy studies. *J. Psychol.*, 1938, **6**, 81-88.
109. MENNINGER, W. C. The treatment of chronic alcohol addiction. *Bull. Menninger Clin.*, 1938, **2**, 101-112.
110. MILES, H. H. W., BARRABEE, E. L., & FINESINGER, J. E. Evaluation of psychotherapy, with a follow-up study of 62 cases of anxiety neurosis. *Psychosom. Med.*, 1951, **13**, 83-105.
111. MORRIS, W. W. Prognostic possibilities of the Rorschach method in metrazol therapy. *Amer. J. Psychiat.*, 1943, **100**, 222-230.
112. MUELLER, M. Prognosis, schizophrenic diseases. *Fortschr. Neurol. Psychiat.*, 1935, **7**, 447-474.
113. MUENCH, G. A. The application of diagnostic psychological methods to counseling and psychotherapy. In M. R. Harrower (Ed.), *Recent advances in diagnostic psychological testing*. Springfield, Ill.: Thomas, 1950. Pp. 53-72.
114. MYERS, C. R. Prediction in clinical psychology. *Canadian J. Psychol.*, 1950, **4**, 97-108.
115. NAGLE, J. M. Two factors in the prognosis of alcoholism. *Psychiat. Quart.*, 1942, **16**, 633-640.
116. O'CONNELL, J. J., & PENROSE, L. S. Tests of psychomotor efficiency in patients treated with metrazol. *J. ment. Sci.*, 1941, **87**, 183-191.
117. OPPENHEIM, S., & BROWER, D. The effects of electric shock therapy as revealed by the Rorschach technique. *Psychiat. Quart.*, 1948, **22** (supplement), 318-325.
118. PACELLA, B. L., PIOTROWSKI, Z., & LEWIS, N. D. C. The effects of electric convulsive therapy on certain personality traits in psychiatric patients. *Amer. J. Psychiat.*, 1947, **104**, 83-91.
119. PEARSON, J. S. Prediction of the response of schizophrenic patients to electro-convulsive therapy. *J. clin. Psychol.*, 1950, **6**, 285-287.
120. PEARSON, J. S., & SWENSON, W. M. A note on extended findings with the MMPI in predicting response to electro-convulsive therapy. *J. clin. Psychol.*, 1951, **7**, 288.
121. PETERS, H. N. Traits related to improved adjustment of psychotics after lobotomy. *J. abnorm. soc. Psychol.*, 1947, **42**, 383-392.

122. PIERS, G. Prognostic observations in insulin treatment of schizophrenia. *Elgin Pap.*, 1941, 4, 34-45.
123. PIOTROWSKI, Z. The prognostic possibilities of the Rorschach method in insulin treatment. *Psychiat. Quart.*, 1938, 12, 679-689.
124. PIOTROWSKI, Z. Rorschach manifestations of improvement in insulin treated schizophrenics. *Psychosom. Med.*, 1939, 1, 508-526.
125. PIOTROWSKI, Z. A simple experimental device for the prediction of outcome of insulin treatment in schizophrenia. *Psychiat. Quart.*, 1940, 14, 267-273.
126. PIOTROWSKI, Z. The Rorschach method as a prognostic aid in the insulin shock treatment of schizophrenics. *Psychiat. Quart.*, 1941, 15, 807-822.
127. PIOTROWSKI, Z. The modifiability of personality as revealed by the Rorschach method: methodological considerations. *Rorschach Res. Exch.*, 1942, 6, 160-167.
128. PIOTROWSKI, Z., & LEWIS, N. D. C. An experimental criterion for the prognostication of the status of schizophrenics after a three-year interval based on Rorschach data. In P. H. Hoch and J. Zubin (Eds.), *Relation of psychological tests to psychiatry*. New York: Grune & Stratton, 1951. Pp. 51-72.
129. RABIN, A. I. Homicide and attempted suicide: A Rorschach study. *Amer. J. Orthopsychiat.*, 1946, 16, 516-524.
130. RABIN, A. I. Effects of electric shock treatment upon some aspects of personality and intellect. *Amer. Psychologist*, 1947, 2, 284. (Abstract)
131. REES, W. L., & JONES, A. M. An evaluation of the Rorschach Test as a prognostic aid in the treatment of schizophrenia by insulin coma therapy, electronarcosis, electroconvulsive therapy and leucotomy. *J. ment. Sci.*, 1951, 97, 681-689.
132. REIMAN, M. G. The prognostic value of mental symptoms in the psychoses. *Stud. Psychol. Psychiat.*, 1935, 3, No. 6. Pp. 40.
133. RENNIE, T. A. C. Analysis of one hundred cases of schizophrenia with recovery. *Arch. Neurol. Psychiat.*, Chicago, 1941, 46, 197-229.
134. RHINE, J. B. ESP and prognosis. *J. Insur. Med.*, 1949, 4, 16-17.
135. RIOCH, M. J. The use of the Rorschach Test in the assessment of change in patients under psychosurgery. *Psychiatry*, 1949, 12, 427-434.
136. ROBERTS, L. K. The predictive significance of the Rorschach Technique in relation to the outcome of psychotherapy. Unpublished doctoral dissertation, Univer. of Iowa, 1950.
137. RUESCH, J., et al. Chronic disease and psychological invalidism. *Psychosom. Med. Mon.*, No. 9, New York, 1946.
138. RUESCH, J., & BOWMAN, K. M. Prolonged post-traumatic syndromes following head injury. *Amer. J. Psychiat.*, 1945, 102, 145-163.
139. RUESCH, J., CHRISTIANSEN, C., PATTERSON, L. C., DEWEES, S., & JACOBSON, A. Psychological invalidism in thyroidectomized patients. *Psychosom. Med.*, 1947, 9, 77-91.
140. SALFIELD, D. J. The usefulness of the Rorschach Test for diagnosis, prognosis and epicrisis, mainly in child guidance treatment. *J. ment. Sci.*, 1951, 97, 84-89.
141. SARASON, S. S. The test situation and the problem of prediction. *J. clin. Psychol.*, 1950, 6, 387-392.
142. SARBIN, T. R. The logic of prediction in psychology. *Psychol. Rev.*, 1944, 51, 210-228.
143. SCHERER, I. W. Correlates of improvement with electro convulsive therapy. *Psychol. Serv. Cent. J.*, 1949, 1, 109-121.

144. SCHERER, I. W. Prognoses and psychological scores in electroconvulsive therapy, psychosurgery, and spontaneous remission. *Amer. J. Psychiat.*, 1951, **107**, 926-931.
145. SCHNACK, G. F., SHAKOW, D., & LIVELY, M. L. Studies in insulin and metrazol therapy: 1. The differential prognostic value of some psychological tests. *J. Pers.*, 1946, **14**, 106-124.
146. SCHNACK, G. F., SHAKOW, D., & LIVELY, M. L. Studies in insulin and metrazol therapy: II. Differential effects on some psychological functions. *J. Pers.*, 1946, **14**, 125-149.
147. SCHOFIELD, W. Research in clinical psychology: 1949. *J. clin. Psychol.*, 1950, **6**, 234-237.
148. SCHOFIELD, W. Changes in responses to the Minnesota Multiphasic Inventory following certain therapies. *Psychol. Monogr.*, 1950, **64**, No. 5 (Whole No. 311).
149. SEITZ, P. F. D. A dynamic factor correlated with the prognosis in paranoid schizophrenia. *Arch. Neurol. Psychiat.*, Chicago, 1951, **65**, 604-606.
150. SEXTON, M. C. The Autokinetic Test: Its value in psychiatric diagnosis and prognosis. *Amer. J. Psychiat.*, 1945, **102**, 399-402.
151. SHASKIN, D. A., CONRAD, D. C., & GRANT, J. D. Prediction of behavior in group psychotherapy from Rorschach protocols (preliminary report). *Group Psychotherapy*, 1950, **3**, 218-230.
152. SIEGEL, M. G. The Rorschach Test as an aid in selecting clients for group therapy and evaluating progress. *Ment. Hyg.*, 1944, **28**, 444-449.
153. SIEGEL, M. G. The use of the Rorschach Test in a treatment program. *Rorschach Res. Exch.*, 1945, **9**, 126-129.
154. SIEGEL, M. G. The Rorschach Test in diagnosis and prognosis. *Family*, 1946, **27**, 102-106.
155. SIEGEL, M. G. The diagnostic and prognostic validity of the Rorschach Test in a child guidance clinic. *Amer. J. Orthopsychiat.*, 1948, **18**, 119-133.
156. SIEGEL, S. M. Personality factors in psychotherapeutic improvement as identified by the Rorschach Test. *Amer. Psychologist*, 1951, **6**, 341-342. (Abstract)
157. SILLMAN, L. R. Chronic alcoholism. *J. nerv. ment. Dis.*, 1948, **107**, 127-149.
158. SIMON, W., & HALES, W. Note on a suicide key in the Minnesota Multiphasic Personality Inventory. *Amer. J. Psychiat.*, 1949, **106**, 222-223.
159. SIMON, B., & HOLT, W. L., JR. The relief of specific psychiatric symptoms by convulsive shock therapy. *Dis. nerv. Syst.*, 1946, **7**, 241-244.
160. SKOTTOWE, I. Shock therapy; a plea for proportion in psychiatry. *Proc. roy. Soc. Med.*, 1939, **32**, 843-852.
161. SLOAN, W. Prediction of extramural adjustment of mental defectives by use of the Rorschach test. *J. consult. Psychol.*, 1948, **12**, 303-309.
162. SNYDER, W. U. A comparison of one unsuccessful with four successful nondirectively counseled cases. *J. consult. Psychol.*, 1947, **11**, 38-42.
163. SSUCHAREWA, G. E., & KOGAN, E. I. Prognosis of schizophrenia of childhood and puberty. *Sovetsk. psikhoneur.*, 1933, **6**, 120-131.
164. STAINBROOK, E. Shock therapy: Psychologic theory and research. *Psychol. Bull.*, 1946, **43**, 21-60.
165. TERRY, G. C., & RENNIE, T. A. C. Analysis of parergasia. *Nerv. ment. Dis. Monogr. Ser.*, No. 64, 1938.
166. TILLOTSON, K., & FLEMING, R. Personality and sociologic factors in the prognosis and treatment of chronic alcoholism. *New Engl. J. Med.*, 1937, **217**, 611-615.

167. THOMAS, J. S. Some clinical examples of "dys-symbolic": its relation to shock therapy. *J. ment. Sci.*, 1940, 86, 100-108.
168. VAN BARK, B., & BARON, S. Neurotic elements in the Rorschach records of psychotics. *Rorschach Res. Exch.*, 1943, 7, 166-168.
169. VOTH, A. C. An experimental study of mental patients through the autokinetic phenomenon. *Amer. J. Psychiat.*, 1947, 103, 793-805.
170. WALL, C. Observations on the behavior of schizophrenic patients undergoing insulin shock therapy. *J. nerv. ment. Dis.*, 1940, 91, 1-8.
171. WECHSLER, D., HALPERN, F., & JAROS, E. Psychometric study of insulin-treated schizophrenics. *Psychiat. Quart.*, 1940, 14, 466-476.
172. WENDER, H. Experiences in group psychotherapy with insulin-treated patients. *Psychiat. Quart.*, 1950, 24, 314-323.
173. WHEELER, E. O., WHITE, P. D., REED, E. W., & COHEN, M. E. Neurocirculatory asthenia (anxiety, neurosis, effort syndrome, neurasthenia). *J. Amer. med. Ass.*, 1950, 142, 878-889.
174. WILLENSON, D. Relation of subtest performance on the Wechsler-Bellevue Intelligence Scale to the effects of psychosurgery. Unpublished master's thesis, Columbia Univer., 1950.
175. WINDLE, C., & HAMWI, V. An exploratory study of the prognostic value of the Complex Reaction Time Test in acute and chronic psychotics. To be published.
176. WITTMAN, PHYLLIS. A scale for measuring prognosis in schizophrenic patients. *Elgin Pap.*, 1941, 4, 22-33.
177. WITTMAN, PHYLLIS. Diagnostic and prognostic significance of the shut-in personality type as a prodromal factor in schizophrenia. *J. clin. Psychol.*, 1948, 4, 211-214.
178. WITTMAN, PHYLLIS, & RUSSELL, J. T. Mental efficiency levels before and after shock therapy. *J. gen. Psychol.*, 1942, 26, 3-17.
179. WITTMAN, PHYLLIS, & STEINBERG, L. Follow-up of an objective evaluation of prognosis in dementia praecox and manic-depressive psychoses. *Elgin Pap.*, 1944, 5, 216-227.
180. ZUBIN, J. Discussion I. In P. H. Hoch and J. Zubin (Eds.), *Relation of psychological tests to psychiatry*. New York: Grune & Stratton, 1951. Pp. 96-104.
181. ZUBIN, J., & THOMPSON, JANE. *Sorting tests in relation to drug therapy in schizophrenia*. New York: New York State Psychiatric Institute, 1941.
182. ZUBIN, J., & WINDLE, C. The prognostic value of the Metonym Test in a follow-up study of psychosurgery patients and their controls. *J. clin. Psychol.*, 1951, 7, 221-223.
183. ZUBIN, J., WINDLE, C., & HAMWI, V. Retrospective evaluation of psychological tests as prognostic instruments in mental disorders. To be published.

Received February 18, 1952.

SOME OBSERVATIONS ON Q TECHNIQUE

WILLIAM STEPHENSON

University of Chicago

With the exception of the present author, all who have given some consideration to R and Q^1 techniques have decided that, after all, they are not different in any important respect. Thomson (32) seems undecided, but Burt (3), Cattell (4), Loevinger (19), Babington-Smith (15) and others argue that the same data can be examined both ways, by R or Q , with the same results. If there are more persons than tests, then *tests* are correlated (R), but if there are more tests than persons, then *persons* are correlated instead (Q). All that distinguishes the two at bottom, therefore, seems to be a matter of convenience. Our own view, first given in 1935 (25), is that such arguments are purely superficial, and in no way represent what is at issue. A little logical analysis indicated at once, so it seemed to us, that wholly distinct principles were involved in R and Q respectively, leading to quite different *systems* (27) as we first named them, that is to different *methodologies* as they are now called. Burt and the present writer, in particular, agreed to differ about these issues (3).

However, the position is now much clearer than it was in 1935, thanks to advances in modern logical analysis and to the influence of Fisher's (9) methodology. It is now certain that not only are R and Q as distinctive as we said they were, but also that the former is based upon many fallacies which are obviated in Q . It is the purpose of the present paper to outline some of these matters.

What is at issue is roughly as follows. When Spearman introduced factor analysis it appeared to offer unusual promise. But serious mistakes were made at its birth. It was linked to interdependency analysis (15) and, consequently, to statistical speculations about "factors of the mind," "unitary factors," "primaries," and the like. Sight was lost altogether of quite a different possibility, that factor analysis might serve, instead, merely as an adjunct to dependency analysis, in which the concern is with psychological experiments and not with statistical interdependencies. We defined Q , originally, merely as an experimental adjunct (26, 27). Similarly, because of wrong conceptions about scientific method, it was protopostulatory to R that psychological theories are the same thing, methodologically, as general propositions, which are

¹ Cattell (4) has added other letters of the alphabet to these two, but his various applications stem, methodologically, from the four *systems* originally defined by Stephenson (27), and, as the present paper will indicate, the additional letters have reference to *variate-designs* and not to *methodologies* of the order we are to discuss.

testable for their "general implications." We now know that this is not only unsound in principle, but that it could do no other than stultify all attempts to explore psychological theories in their own right, with respect to the correct logic of singular propositions derived from the theories (14, 23). It is essentially this latter logic of scientific method that *Q* achieves, and its promise for psychology, consequently, is not less than we claimed for it originally. In reintroducing *Q* methodology, therefore, we have the double task of adjusting many far-reaching mistakes of the early factorists, and of giving a foretaste of the wide applications of factor analysis as an adjunct to experimental method.

EARLY FORMULATIONS

The applications of factor analysis that are best known, of course, are those with which Spearman began his work, now called *R* technique, in which the individual differences provided by *tests* or the like are correlated and factored. *R* technique is a technical way of studying individual differences for attributes of persons. *Q* technique was defined in 1935 (27) as a *system* or *methodology*, radically different from that for *R*. Clearly others, including Beebe-Center (1) and Burt (3), had correlated persons long before Thomson (32) and Stephenson (26) drew attention to these systematic possibilities, but the grasp of this matter in its widest methodological respects was overlooked prior to our early papers (25, 26), and has been missed up to now by many who have sought to expound *Q* technique, including Burt (3), Cattell (4), Babington-Smith (15), and Loevinger (19).

We defined *R* and *Q* as two independent systems, the one concerned with individual differences by postulation, and the other not. Burt (3) offered a proof that the two were related, but his *reciprocity principle* merely dealt with any one matrix, doubly centered, and with covariance analysis, and therefore touched upon neither *R* nor *Q*, much less any conceivable relation between them (32). By postulation, *R* and *Q* always involve *two* quite different, and singly centered, tables of correlations, each subserved by its own distinctive quantitative and qualitative principles. It is therefore a mistake to argue as though all that is involved is a single matrix of data which, when correlated down the rows is *R*, and along the columns is *Q*. Quantitative principles for *Q* were described in 1935, and were perhaps never taken seriously enough by our critics. They may be summarized briefly as follows: (a) The populations are statements, traits, or the like; (b) variates refer to operations of a single person, or about him, in one interactional setting; (c) the transitory postulate (16) has reference to intra-individual differences of "significance"; (d) variates may interact; (e) scores are approximately

normal and standardized, as in product-moment correlational theory generally; (f) all the important information for each array is contained in its variation (no information is contained in the variate means); (g) the operations of, or about, a person are all subject to the principle of randomization (9); and (h) the concern is with dependency analysis.

The quantitative principles for *R* concern persons as populations, attributes as variates, and the transitory postulate is made to work in terms of individual differences; the variates (tests, etc.) in *R* do not interact—operations are all subject to the “rule of the single variable”; nor do persons interact; and the concern is with *interdependency*, and not with *dependency* forms of analysis. The qualitative principles are also very different for *R* and *Q* techniques. Those for *R* technique have been based upon an essentially inductive methodology, and upon the study of psychological theories for their “general implications.” In *Q* methodology, on the contrary, the concern is with a postulatory-dependency methodology, and with psychological theories as growing-points for singular testable propositions. We believe that the former methodology has had seriously restrictive effects upon psychology, whereas the latter is in keeping with the modern logic of scientific method, and offers almost unlimited scope for an experimental approach to psychology.

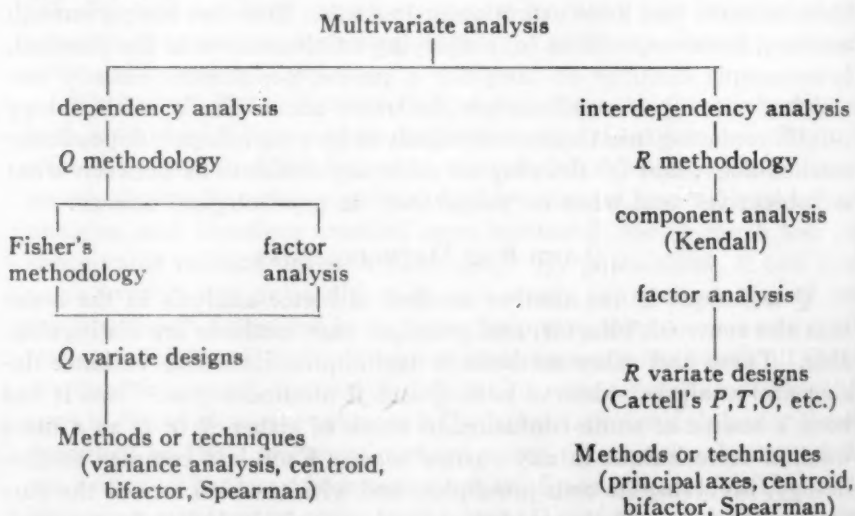
We must introduce, therefore, a number of basic formulations which have hitherto had little attention in journals. They are simple enough matters, however, such as (a) employing an alternative to the classical, large-sample doctrine of sampling a parent population, namely one which draws upon small-sample doctrines and Fisher's methodology (9), (b) replacing interdependency analysis by a postulatory-dependency methodology, and (c) drawing no arbitrary distinctions between what is “objective” and what is “subjective” in psychological science.

Q AND R AS METHODOLOGIES

Q technique is not another method of factor analysis in the sense that the centroid, bifactor, and principal axes methods are distinguishable. These and other methods or techniques, including variance design and analysis, subserve both *Q* and *R* methodologies. Thus it has been a source of some confusion to think of either *R* or *Q* as a mere method or technique in any narrow sense. Each is a complex methodology, involving its own principles, and within which any of the current methods or techniques of statistical analysis, including factor, variance, probability analysis, and the like, may be applied. But, up to now factor analysis has been regarded widely as a matter of interdependency analysis (3, 15); almost everyone believes that the concern is with the

discovery of factors as unitaries (32) or primaries (33). An inductive methodology is at issue in which factors have first to be found, and afterwards interpreted (33). Our own standpoint in *Q* methodology is very different from this: we postulate hypotheses, explanations, and interpretations at the outset in relation to a psychological theory; a set of propositions is then asserted which is put to an empirical test. To the latter end (a) structured *Q* samples are composed, where possible, which entail the independencies of the theory at issue, or, if unstructured samples are used, the independencies are implied, and (b) variate-designs are employed, with the object of bringing dependencies to light, that is, such as put the propositions to test. *Q* technique, in a narrow sense, consists of experimenting with these two kinds of designs (a) and (b); factors represent dependent effects and, in certain circumstances, the possibility of lawfulness and underlying tendencies. The effects are tested for significance by any statistical methods that may be appropriate to the situation, variance analysis (Fisher) and centroid factor procedures (Thurstone) being the most immediately useful. Thus, *Q* methodology is concerned with testing concrete, singular, propositional sets.

Following Kendall's example (15) it may be of some assistance to offer a kind of genealogical tree for the situation vis-a-vis *R* and *Q* as methodologies:



The cleavage into interdependency and dependency forms of analysis is indicated. In the former case no variate is picked out for prior or any special regard (15, 34), whereas in the latter the concern is with in-

dependent and dependent variables. The variate designs can be legion, far exceeding any number of letters of the alphabet; the various so-called systems of Cattell (4) have the status of such variate designs. The joining of Fisher's methodology to factor analysis is indicated for *Q* methodology. What is involved in dependency analysis has always been regarded as the essence of experimental method, namely, that independent variables or independencies should be specified, and, as a result of operations in relation to them, dependent effects should be obtained. Usually only one dependent effect is studied at a time, but in Fisher's methodology many may be so studied contemporaneously. We seek to achieve such a methodology, then, for *Q*. The greatest single aid to this end is in the use of structured *Q* samples, a matter to which we shall now attend.

ON STRUCTURED SAMPLES

Large sampling conditions are at issue in *R* and small sampling conditions in *Q*. First a distinction is drawn between populations and statistical universes (20). A sample of 200 children in *R* is a population sample, but the scores of these children on a test constitute a statistical distribution. For any one person-population there can be innumerable statistical universes. When large sampling conditions are strictly adhered to, area, stratified, biased, controlled and similar devices are employed in order to reach representative sampling or to allow for departures from it. Our particular innovation has reference to population samples: we propose that these samples can be structured with respect to specifiable independencies, for balanced block or other Fisherian designs (9). The general formulation is as follows:

If there can be defined certain independencies *A*, *B*, *C*, with "levels" *a*, *b*, *c* . . . respectively, then, without replication, but with a balanced block design, there are $a \times b \times c$ combinations, one "level" at a time for each independency. If these are replicated *m* times, there will be $m \times abc$ such combinations. We

TABLE 1
FACTORIAL DESIGN FOR A SAMPLE OF PERSONS

<i>Independencies</i>	<i>Levels</i>	<i>No.</i>	<i>Degrees of freedom</i>
W. Age	(a) 15-20 years (b) 20-30 years	2	1
X. Socioeconomic status	(c) A (d) B (e) CDE	3	2
Y. Educational status	(f) University (g) High School	2	1
Z. Habitat	(h) Rural (i) City	2	1

define a structured sample in this way, to consist of a set of *mabc* cases in balanced design.

It is possible, of course, to use other designs as well, such as latin squares, or confounded designs. In *R* methodology a typical design for population samples could be that shown in Table 1. For this design there are $2 \times 3 \times 2 \times 2 = 24$ combinations of the levels, one at a time from each independency, namely, the following:

<i>a a a a</i>	<i>a a a a</i>	<i>a a a a</i>	<i>b b b b</i>	<i>b b b b</i>	<i>b b b b</i>
<i>c c c c</i>	<i>d d d d</i>	<i>e e e e</i>	<i>c c c c</i>	<i>d d d d</i>	<i>e e e e</i>
<i>f f g g</i>	<i>f f g g</i>	<i>f f g g</i>	<i>f f g g</i>	<i>f f g g</i>	<i>f f g g</i>
<i>h i h i</i>	<i>h i h i</i>	<i>h i h i</i>	<i>h i h i</i>	<i>h i h i</i>	<i>h i h i</i>

If the concern is with American white men, 24 men can be chosen to cover these 24 possibilities. Or, if a larger sample is required, each combination or "cell" can be replicated as many times as necessary. For 10 replications the sample would be $n = 240$. The design as such was first employed in psychology by Crutchfield (5). In his case, however, the concern was to define an experimental situation in which to place an unstructured set of rats randomly, whereas our concern is to define a structured population sample. Clearly, statistical universes are not involved up to this point, but if structured samples of the above kind can be suitably quantified (e.g., if, in the above case, each of the persons of a sample gains a score on a particular mental test), the methods of variance analysis are at once applicable to the data, for the following apportioning of the variance (for Table 1):

	<i>df</i>
ΣW	1
ΣX	2
ΣY	1
ΣZ	1
Σ (interactions)	18
Σ (replication)	24 ($m - 1$)

The same procedure is followed in *Q* methodology, except that statements, traits, pictures, and the like constitute the populations, and not persons, and *Q* sortings give rise to the statistical distributions, and not individual differences. The first examples of the kind we employed concerned attitudes towards scholastic subjects (28), and later, Jung's theory of personality (29). The latter provides as good an example as any. Jung's theory specified three main "effects," *X*, the "attitudes" of introversion-extroversion; *Y*, the "mechanisms" (conscious-unconscious); and *Z*, the "functions," (namely thinking, feeling, sensation, and intuition).

The design for these is shown in Table 2.

TABLE 2
 FACTORIAL DESIGN FOR A SAMPLE OF JUNGIAN STATEMENTS

Independencies	Levels		No.	Degrees of freedom
X. Attitudes	(a) introversion	(b) extroversion	2	1
Y. Mechanisms	(c) conscious	(d) unconscious	2	1
Z. Functions	(e) thinking	(f) feeling	4	3
	(g) sensation	(h) intuition		

This leads to $2 \times 2 \times 4 = 16$ combinations of the independencies, one level at a time. In order to clothe it with statements, we merely take assertions by Jung (13) which comport with these combinations, one statement for each combination (or as many as there are replications of the design). Thus, Jung's statement "ready to sink a battleship or to amputate a leg" fits the combination *bcf* (according to Jung's theory); "quietly sensual" is *adg*. A set of 16 such statements is readily found in Jung's work to cover the 16 possibilities of the theory. But we can replicate, taking say five statements for each "cell," and composing in this way a sample of size $n = 80$. If required the replications can be drawn at random from pools of the statements made by Jung (or on theoretical grounds so attributable). Theoretically any number of such samples can be composed for the given design, and any one is in principle as representative of the theory as any other. In this way many difficulties about sampling conditions in *Q* technique have been obviated, such as concern lack of independency and contingencies (15).

The *Q* sample is merely a population sample. But it embodies the independencies of a theory, and we have found the procedure very widely applicable to samples in social psychology, clinical psychology,

TABLE 3
 FACTORIAL DESIGN FOR A SAMPLE OF RORSCHACH STATEMENTS

Independencies	Levels			No.	Degrees of freedom
X. Control	(a) outer	(b) inner	(c) repressive (constrictive)	3	2
Y. Adjustment	(d) systematized anxiety	(e) unsystematized anxiety	(f) balanced	3	2
Z. Erlebnistype	(g) introvertive	(h) extrovertive		2	1

aesthetics, and for personality theory in general. Thus, in order to represent Rorschach's theory for his projective technique, the design in Table 3 is appropriate. The design leads to $3 \times 3 \times 2 = 18$ combinations without replication, and may be covered by Rorschach indicators of the kind $(afh) = FC > CF + C$, or by statements of the inferential kind that have reference to such indicators, e.g., "sublimating" (bfg). We can also represent in this way conclusions that have been drawn from previous studies with respect to presumed independencies.

THE POSTULATORY-DEPENDENCY METHODOLOGY

The structured samples arrived at in the above manner are not testable propositions in terms of which we seek to prove or disprove a theory or hypothesis by validating the postulated independencies for their "general implications." That is, there is no question of *proving* that the statements in Q , any more than the persons in R , are as they are asserted to be "on the average," or "in general," or as "indicated from the study of individual differences." The Q samples are used, instead, for singular tests of propositions that the theory suggests, or that are deduced from it.

The methodology is not restricted to broad theories, but it is easiest to discuss it in terms of theories such as Spearman propounded about noegenesis (24), Jung about personality, Freud about the unconscious, or Rogers about the self. These would probably be regarded as unscientific per se or *in esse* from the R methodology standpoint; or, if a so-called scientific approach were made to them along R technique lines, as was attempted for the theories of Jung and Spearman, it would be assumed that general propositions are at issue, to be examined for their "general implications." This confuses a rational theory, however, with a general or universal proposition. Theories should be regarded, instead, as growing-points for singular propositions (14, 23). Thus, many attempts have been made in the past to measure introversion-extroversion, based on the supposition that all persons are introverts or extroverts in some degree habitually. The proposition is characteristic of all R technique studies, since everyone is supposed to have all attributes in some degree. The Guilfords (10) and others, as is well known, found several factors instead of one, such as (they supposed) Jung's theory adumbrated: but each R factor has the same postulation at its roots—all persons have each in some degree. This indeed seems inescapable and self-evident.

However, such general implications are nowhere necessarily involved in the theory. The concern, instead, can be with singular propositions, of the following kind.

P.1: *This particular person X is either introvert or extrovert habitually, or neither.*

This we can at once test along Q lines, in terms of the *theory* of introversion-extroversion, without reference to any person other than X, and without the use of any norms or standard scales, and without operational reference to any individual differences. We merely invite X to offer a self-description of himself, as he conceives himself to be habitually, in terms of a Q sort, for a structured sample based on Table 2. In a particular case, for example, we had available a sample of 160 statements, i.e., for 10 replications of the design of Table 2. Our landlady was invited to give a self-description,² for the following frequency distribution of scores:

	Most characteristic of me habitually						Least characteristic of me habitually			
Score	8	7	6	5	4	3	2	1	0	
Frequency	8	12	20	24	32	24	20	12	8	
	<i>n</i> = 160									

The scores provided in this way in the respective "cells" of the design are shown in Table 4.

The analysis in terms of "expectancies" proceeds as shown in Table 5. The *F* test shows that only the first of these sums-of-squares is significant. The data then clearly indicate that it is (b), i.e., the introversion level, which has been given the significantly large score. That is, our landlady operates in such a way as to suggest that, in terms of Jung's theory, she is introvertive. None of the other effects is significant in her case.

² This consists of arraying the statements of the sample in an order from those most characteristic of her, *in her own view*, to those least characteristic. The statements are typed on cards, one statement to a card, and the subject first reads them through in order to grasp their import. They are then shuffled, and she proceeds to the "Q sort," as we call it: usually the cards are first divided roughly into three piles, one for those that characterize her positively, and one for those that could scarcely do so under any conditions (she believes), with the doubtful or neutral ones in between. The three piles are then further teased apart, working from the two extremes, until she has provided the required forced frequency distribution. Thus 8 of the cards, which she decides are most characteristic of all gain 8 marks, the next 11 most characteristic, 7 marks . . . and so on. For ease of application we rarely employ as many items as 160 for a single Q sort. In the present case the 160 was divided into two samples of 80 each, the Q sorting being repeated for each in turn and the results combined. Sometimes a final re-sort is undertaken in such a case, in which the two sets of cards, previously arrayed by the subject in their descriptive order, are now placed in parallel before her, so that she can undertake such minor adjustments in position as she may deem necessary when all 160 so confront her. The "forced frequency" distribution has many practical advantages and is by no means as arbitrary as it may appear to be at first sight. We use only quasi-normal distributions, platykurtic in shape, for certain theoretical reasons, and we find it best to have the range of scores from 0 to 12, rather than 0 to 8 as in the present example.

TABLE 4

RESULTS OF Q SORT FOR A STRUCTURED SAMPLE OF 160 STATEMENTS*

Cells	a	a	a	a	a	a	a	a	a	b	b	b	b	b	b	b	b
	c	c	c	c	d	d	d	d	e	e	e	e	d	d	d	d	d
	e	f	g	h	e	f	g	h	e	f	g	h	e	f	g	h	h
Scores	0	5	1	6	5	4	5	4	2	6	5	6	7	8	6	4	
	6	4	3	4	4	4	4	3	8	5	4	6	6	7	5	6	
	6	3	2	3	3	2	3	2	7	3	4	4	5	4	4	6	
	5	2	0	0	2	0	1	4	6	2	3	3	4	4	3	6	
	4	2	0	0	7	1	8	1	4	8	2	8	3	2	5	3	
	7	6	1	6	4	4	3	6	3	6	2	7	2	7	3	3	
	4	4	4	5	1	3	0	5	5	5	7	5	2	6	5	5	
	4	3	4	1	1	3	4	5	3	4	1	5	5	3	2	4	
	3	5	4	8	5	1	7	1	5	7	4	2	2	2	8	5	
	2	0	1	5	8	2	2	4	4	3	6	6	7	5	6	7	

* Based upon the design shown in Table 2.

The replication variances are tested for *homogeneity* (2) in the usual way, as a consequence of passing which we may suppose that the experimental subject has operated with the statements according to the principle of randomization (Fisher, 9), so that contingencies of any obvious kind are not at issue. Clearly anyone can take part in such an experiment, and the conditions can be made as specific as we care to make them: Mrs. X can describe herself (a) as she was ten years ago, (b) at her happiest, (c) as she feels herself to be at a party, and so on. Every Q sort, for anyone, can be analyzed in the above manner, prior to any factor analysis of a number of such variates (whether for different persons, or for one-and-the-same person). Facts can be accumulated in

TABLE 5

RESULTS OF ANALYSIS OF VARIANCE FOR THE DATA FROM THE Q SORT SHOWN IN TABLE 4

	Source of Variance	Sums of Squares	df	"Expectancy"	F
ΣW.	(Between Attitudes)	70.22	1	70.22	18.14
ΣX.	(Between Mechanisms)	0.22	1	0.22	0.05
ΣY.	(Between Functions)	15.35	3	5.12	1.32
	(Interactions)	36.61	10	3.66	0.94
	(Replication)	557.50	144	3.87	—
	Total	680.00			

this way which Jung would presumably explain according to his theory (but of course alternative theories would be permissible). And if one didn't trust self-appraisals, observers could offer descriptions of Mrs. X from the "outside" standpoint, preferably as observed under role-playing conditions.

We do not suppose, of course, that any theory is proved or disproved by way of a few such singular studies. Nor, if thousands of such singular propositions are tested satisfactorily about a theory, is the theory necessarily acceptable on that account. Valuable theories point the way to interesting propositions. Thus, instead of examining Jung's theory for its "general implications" our concern nowadays would be with many-propositions which we can test under singular conditions, and which were never so testable hitherto, such as stem from propositions of the following kind:

P.2: Extroverts X, Y . . . have *insights* into another extrovert W, that they do not have about introverts A, B. . . .

P.3: Sophisticated parents of family T are more individuated than their children.

P.4: Phantasy is a "bridge" between X's claims of introversion and extroversion.

P.5: Extrovert X has a certain "repugnance, fear, or silent scorn" for introversion and an introvert Y has the same for extroversion.

Experimental work is undertaken about each of these, and others of the kind, along Q lines, without reference anywhere to the supposed basic principles as testable propositions. That is, the theory is postulated: it is what it leads to that matters, in concrete singular situations.

This may seem a novel way of looking at scientific theory, yet something of the kind is clearly evident in the hypothetico-deductive methodology, with which the postulatory-dependency procedures under discussion have some affinities. We do not seek, however, to fashion any rigid hypothetico-deductive systems at this juncture. But it is certain that the right kind of theories can give rise to innumerable propositions, and that their importance lies rather in the discoveries they make possible than in any "general implications" they may have, of the kind that R methodology seeks to study.

It is therefore our position that two related mistakes were made at the outset of factor analysis: one was to regard theories as general propositions, and the other to regard factor analysis as a matter of interdependency analysis. Spearman, the founder of factor analysis, had a psychological theory in mind at the outset, namely, that of *Noogenesis* (24). According to this, all *new* experiences came by way of the noetic principles, and all *learned* experiences were regarded as anoetic, based on many more specific principles. The famous *Theory of Two*

Factors, therefore, was merely a mathematical model for the psychological theory of noesis: g represented noesis, and all degrees of s , anoesis. This was examined for its "general implications," and the consequence has been the complete disregard of the noetic theory in particular, and of all psychological theory about the abilities in general. Instead of theories about intelligence, purely logical metatheories have been promulgated, such as Burt's *Four-Factor Theory* (3), or Thurstone's *Multiple-Factor Theory* (33), or Thomson's *Sampling Theory* (32). We now know that the theory of noesis should have been regarded merely as a guide for dependent forms of experimenting, in which the theory is at issue for its singular propositions, much as we have suggested for Jung's theory above. But the early factorists were fascinated by the logic of interdependency analysis, with the result that psychological theory has become conspicuously absent from R methodology. Psychological attributes are still being studied for their interdependencies and "general implications," and the search still goes on for unitary and primary factors. It is significant to observe that psychological theories cannot be represented readily in structured R samples, whereas their scope in Q samples seems unlimited. Such is the measure of the difference between R and Q , which our critics have long believed to be merely two sides of the same coin.

Unlike the simple proposition P.1 above, those at P.2 to P.5 may require several variates for their solution, and not just one. The several constitute a variate-design, and one of the arts of experimenting in Q technique consists of designing suitable variates which will serve to put propositions in a testable form. A design for P.2 has already been reported (31): it was shown that members of a class who regarded themselves as extroverts had couplet factors with the experimenter, who regarded himself also as extrovert, whereas the members of the class who maintained that they were introverts did not provide such couplets. A design for the before-during-after-treatment sequence of therapy was employed by Hartley (11) which Mowrer wishes to distinguish from Q , calling it O technique. It is, instead, merely a particular Q variate design, of which hundreds already have been suggested or used.

These variates represent the *dependencies*, as do factors, which subsume a number of variates. Obviously it is possible to correlate such variates, and to factor the correlations: the variance analysis can thereupon be applied to *factor arrays* instead of to the original variates. In this sense factor analysis could be merely a subsumptive device, permitting us to replace many single variates by one or more factor variates, each of which can be examined for its dependencies along Fisherian lines. However, this is not the only function we require factor analysis to serve in Q methodology.

Meanwhile we may suppose that the typical *Q* study now recognizes (a) postulated independencies of a theory, whether made explicit in structured samples, or left implicit in unstructured ones, (b) theorems, hypotheses, or propositional sets having reference to the consequences of the theory, (c) variate designs which serve to put these sets to empirical test, (d) the variates themselves, and the factors which they subsume, representing dependent effects. Under certain conditions, (e) it is expected that the factors will point to underlying tendencies, and to lawfulness of a kind not asserted at the outset in the original theory. Factors can be interpreted, usually, in terms of the postulated theory. But such interpretations are not the essential objective of *Q* studies, which should lead to more abstract explanations which are not incompatible with these lower level ones, but which point to lawfulness and to the possibilities of general principles not perhaps previously anticipated. Thus, in studies of the self-notions of a person *X* we may reach factors α , β , γ These will be particular to *X*, no other person in the world having them if they are based on a *Q* sample of his own idiosyncratic self-reflections. But α , β , γ . . . may be in dynamic relationship, such as may suggest that if any one factor is altered (by therapy or the like) the others will change *pari passu*. It is at this more abstract level that important general principles are to be expected in *Q* studies in general, as in all scientific work.

However, we do not in fact need to structure all *Q* samples explicitly. When statements are taken which have reference, say, to psychoanalytic theory, the independencies are implied in the statements. The statements that we may take randomly from Jung's work on personality imply his theory, whether they are composed into a structured sample or not. In general, therefore, we may pursue a dependency form of factor analysis for unstructured samples, in so far as we may assert that any factors may be given an explanation in terms of the implied theory. In point of fact many propositions can be asserted beforehand for unstructured samples, just as if we had, in fact, a structured sample available.

Two forms of dependency analysis are employed in *Q* studies, that of variance design already introduced, and dependency factor analysis. Only brief reference can be made to the latter in this paper. It will be apparent to the statistician, however, that the same results can be reached, for a given matrix of data, by the variance analysis of balanced block *Q* arrays, and by their factor analysis. Effects *X*, *Y*, *Z* in variance design are orthogonal, and significant interactions are merely additional to these, but not postulated at the outset. Precisely the same effects can be postulated as orthogonal *factors*, and any not postulated can be "discovered" as an additional factor or factors. The permissiveness of

the centroid method then permits us to factor our data, and to rotate until the same solution is reached as can be provided by the variance analysis of each Q array. The purpose of the latter analysis is mainly to permit us to classify variates which are alike with respect to specifiable effects, within stated fiducial limits. We employ factor analysis for this same purpose. The object of dependency factor analysis, it may be said, is to look at a matrix of data to determine what Fisherian design can fit it, whether one was postulated in the Q sample originally or not. Simple structure in Thurstone's sense (33) is merely such a design, albeit a confounded one when correlated factors are involved. But these are matters for another occasion.

We would like to add that the procedures we are describing apply to so-called subjective appraisals made by any subject X , or to objective ones, made about X by observers. Long ago an example was provided of the use of Q technique in the study of performance (30) as objectively regarded. The studies apply, no less, to any "single case": logically they are all best so begun. Generalizations concern abstract, higher order levels of interpretation, and are not matters having reference to "large numbers of cases" as such. Many of our most important Q studies involve only a single experimental subject, who may, in principle, be *anyone*.

CONCLUSION

The new methods we are discussing are essentially simple and straightforward in practice. It is rarely necessary in Q technique to deal with more than three or four factors and their combinations in order to analyze data, and no analysis need take more than a matter of a few days to complete. Nor need any variance analysis of data be oppressive: indeed it is rarely essential to resort to it in complex Q studies, since factor analysis provides the same results more economically.

The methodology has very wide applications, as we asserted for it at the outset (25). Self-psychology can now have its propositions directly represented (11, 22). Many interesting studies in clinical psychology have already been undertaken (Hartley (11), Heine (12), Pemberton (21), Fiedler (7, 8)), and also in socio-educational theory (Ebermann, 6). With regard to social psychology we study attitudes now as theoretical issues, and not as "things" to be measured (18) which have this-or-that attributes (17). The projective techniques can now have their theoretical formulations represented as independencies in Q samples. Type psychologies can be more sensibly studied, freed from the erroneous beliefs engendered about them by the assumption of general propositions. Individual differences are seen as merely of technological interest, and in no way necessary as *postulates* to any psychological theory.

The methodology clearly favors a frank acceptance of theories in psychology. Nowhere in it is there any measurement-at-any-price, for no-one-knows-what. The proposal to regard factor analysis as a matter within the domain of dependency analysis is important; in this way factor analysis becomes little more than a complicated kind of *t* test, and nowhere is there any search for, or belief in, primary factors, unitaries, or the like. The methods we describe clearly open to our operational regard much that has hitherto been called "subjective": the only distinctions we can accept between what is subjective and what objective rests upon whether reliable operations are possible or not (Zilsel, 35). Finally, the pertinency of modern logic and analytical philosophy is apparent in our formulations. The distinction between propositions examined for their "general implications," and theories investigated with respect to singular propositions is a case in point, but the quantitative principles upon which *Q* technique is based are such as logical analysis supports. It is because these matters were neglected that *R* and *Q* were so long needlessly confounded.

REFERENCES

1. BEEBE-CENTER, J. G. *The psychology of pleasantness and unpleasantness*. New York: Van Nostrand, 1932.
2. BROWNLIE, K. A. *Industrial experimentation*. London: H.M.S.O., 1949.
3. BURT, C. Correlations between persons. *Brit. J. Psychol.*, 1937, 28, 59-96.
4. CATTELL, R. B. On the disuse and misuse of *P*, *Q*, *Qs*, and *O* techniques in clinical psychology. *J. clin. Psychol.*, 1951, 7, 203.
5. CRUTCHFIELD, R. S. Efficient factorial design and analysis of variance illustrated in psychological experiments. *J. Psychol.*, 1938, 5, 339-346.
6. EBERMANN, P. The application of *Q*-technique to one aspect of teaching competency. Unpublished Ph.D. thesis. Univer. of Chicago, 1951.
7. FIEDLER, F. E. The concept of an ideal therapeutic relationship. *J. consult. Psychol.*, 1950, 14, 239-249.
8. FIEDLER, F. E. A method of objective quantification of certain counter-transference attitudes. *J. clin. Psychol.*, 1951, 8, 101-107.
9. FISHER, R. A. *The design of experiments*. London: Oliver and Boyd, 1942.
10. GUILFORD, J. P., & GUILFORD, R. B. An analysis of the factors present in a typical test of introversion-extroversion. *J. abnorm. soc. Psychol.*, 1934, 28, 377-399.
11. HARTLEY, M. A *Q*-technique study of changes in the self-concept during psychotherapy. Unpublished Ph.D. thesis, Univer. of Chicago, 1951.
12. HEINE, R. An investigation of the relationship between changes and the responsible factors as seen by clients following treatment by psychotherapists of the psychoanalytic, Adlerian, and non-directive schools. Unpublished Ph.D. thesis, Univer. of Chicago, 1950.
13. JUNG, C. G. *Psychological types*. New York: Harcourt, Brace, 1924.
14. KAUFMANN, F. *Methodology of the social sciences*. London: Oxford Univer. Press, 1944.
15. KENDALL, M. G., & BABINGTON-SMITH, B. Factor analysis. *J. roy. statist. Soc. (B)*, (London), 1950, 12, No. 1.

16. KEYNES, J. M. *A treatise on probability*. London: Macmillan, 1921.
17. KRAMER, B. Dimensions of prejudice. *J. Psychol.*, 1949, 27, 389-451.
18. KRECH, D., & CRUTCHFIELD, R. *Theory and problems of social psychology*. New York: McGraw-Hill, 1948.
19. LOEVINGER, J. Intelligence. In H. Helson (Ed.), *Theoretical foundations of psychology*. New York: Van Nostrand, 1951. Pp. 557-601.
20. PEATMAN, J. G. *Descriptive and sampling statistics*. New York: Harper, 1947.
21. PEMBERTON, W. A. Factors involved in clinical judgments made from projective and non-projective behavior samples. Unpublished Ph.D. Thesis. Univer. of Chicago, 1951.
22. ROGERS, C. *Client-centered therapy*. New York: Houghton, Mifflin, 1951.
23. SCHLICK, M. Facts and propositions. *Analysis*, 1945, 2, 65-70.
24. SPEARMAN, C. *The nature of intelligence and the principles of cognition*. New York: Macmillan, 1925.
25. STEPHENSON, W. Technique of factor analysis. *Nature*, 1935, 136, 297.
26. STEPHENSON, W. Correlating persons instead of tests. *Charact. & Pers.*, 1935, 4, 17-24.
27. STEPHENSON, W. The foundations of psychometry; four factor systems. *Psychometrika*, 1936, 1, 195-209.
28. STEPHENSON, W. A new application of correlation to averages. *Brit. J. educ. Psychol.*, 1936, 6, 43-57.
29. STEPHENSON, W. Methodological consideration of Jung's typology. *J. ment. Sci.*, 1939, 85, 185-205.
30. STEPHENSON, W. Two contributions to the theory of mental tests, II—A statistical regard of performance. *Brit. J. Psychol.*, 1940, 30, 230-247.
31. STEPHENSON, W. The significance of Q-technique for the study of personality. In M. L. Reymert (Ed.), *Feelings and emotions: The Mooseheart symposium*. New York: McGraw-Hill, 1950. Pp. 552-570.
32. THOMSON, G. H. *The factor analysis of human ability*. London: Univer. of London Press, 1939.
33. THURSTONE, L. L. *Multiple-factor analysis*. Chicago: Univer. of Chicago Press, 1947.
34. WOODWORTH, R. S. *Experimental psychology*. New York: Methuen, 1939.
35. ZILSEL, E. Logical empirical analysis. *Int. Encycl. Unif. Sci.*, Vol. II, No. 8. Chicago: Univer. of Chicago Press, 1941.

Received January 6, 1952.

THE THREE BASIC FACTOR-ANALYTIC RESEARCH DESIGNS.—THEIR INTERRELATIONS AND DERIVATIVES

RAYMOND B. CATTELL

University of Illinois

Factor analysis began with the correlation of tests measured on populations of persons, but other arrangements have since been stumbled upon, or deliberately thought out for special purposes, from time to time. In 1946 the present writer formulated the *covariation chart* (9) which integrated in a single conception the accumulation of existing usages and revealed certain new, logically-possible designs of factor analysis.

The purpose of this review of current practice is to call attention to some possible misconceptions and show new directions of practical usefulness. The first effect of the examination of logical possibilities in the covariation chart was to provoke a realization that up to the time of that analysis only a small corner of the universe of effective factor-analytic designs had become actively inhabited by researchers. The "chart" thus proffered powerful new covariation tools, especially in relation to the multivariate problems of clinical psychology, sociology, and physiology, which, except for a few recent examples in *P* and *Q* techniques, still need illustration. It is proposed here to develop those theorems into more explicit practical corollaries for experimental work and to investigate the true interrelations and precise limitations of the various techniques. For in some recent approaches, e.g., the use of *Q* technique by Stephenson (31), it would seem that there is some loss of perspective on methodological relationships.

BASIC REFERENTS IN COVARIATION INVESTIGATIONS

All scientific method deals with observations of covariation, but factor analysis covers that half of the methodological realm which has to do with simultaneous variation in many variables, not the univariate variation of so-called "controlled" classical experiment (16). In either region a single act of measurement has five essential referents or signatures, as follows:

1. A defined set of *circumstances*, time, place, etc., in which the attribute (reaction, trait, operation) is observed. In psychology this is the "stimulus situation."
2. The *attribute* itself, which is defined by an operation of observing or measuring certain things. In psychology this is the "response."
3. An *object*, usually in psychology an organism, to which the attribute is referred.

4. If the observation is to be quantified there is reference also to a *scale* or unit by which the measurement is to be rendered numerical.

5. An *observer*, or, in behavioral data, a set of observers capable of mutually confirmatory evidence.

Although these are exhaustive of the essential signatures for an act of measurement, each of the five is susceptible to some subdivision into subparameters. For example, the stimulus situation has many dimensions besides those of time and place required to define it, and therefore to define the measurement. However, we do not normally expressly define all of these, but merely give sufficient direction to fix and reproduce relevant *circumstances or conditions*. It is unfortunate for clarity that the term "test" is often regarded as defining both stimulus situation and response, whereas it defines wholly only the response measured. The definition of stimulus situation or occasion must be regarded as an additional referent, in which the test material is only a part. For example, an intelligence test still needs definition of the stimulus conditions in which it is given.

For the great majority of psychological experiments in which factor analysis is used we can reduce the essential referents from five to three, namely: circumstances, persons, and attributes, wherein the attribute is an operation of measurement which includes reference to that part of the stimulus situation (5) which remains fixed, and the circumstance (or "occasion") referent is restricted to whatever in the situation varies from occasion to occasion. This reduction to three referents is convenient for initial presentation of the main issues, but we shall include all five later.

If these three primary definers of a psychological observation are arranged as three distinct series (geometrically as axes) we get the covariation chart, as shown in Fig. 1, within which all possibilities of correlation for factor analytic work should be contained (except for the special extensions of the two remaining parameters).

Thus the commonest correlation is on a series of persons, each measured on two attributes, and representable in the chart by two parallel lines, as shown in the channel labelled "*R* technique," starting from two attributes ("tests") j_6 and j_7 . Incidentally, it should be kept in mind that mathematically speaking these axes are not continuous or ordered but represent discrete series, i.e., populations of individual reference points (tests, persons, occasions) having any order in which the sampling happens to present them.

Any pair of parallel lines drawn within the parallelepiped of the covariation chart will represent a possible correlation, for it will indicate measurements of one character, made in two different forms upon a

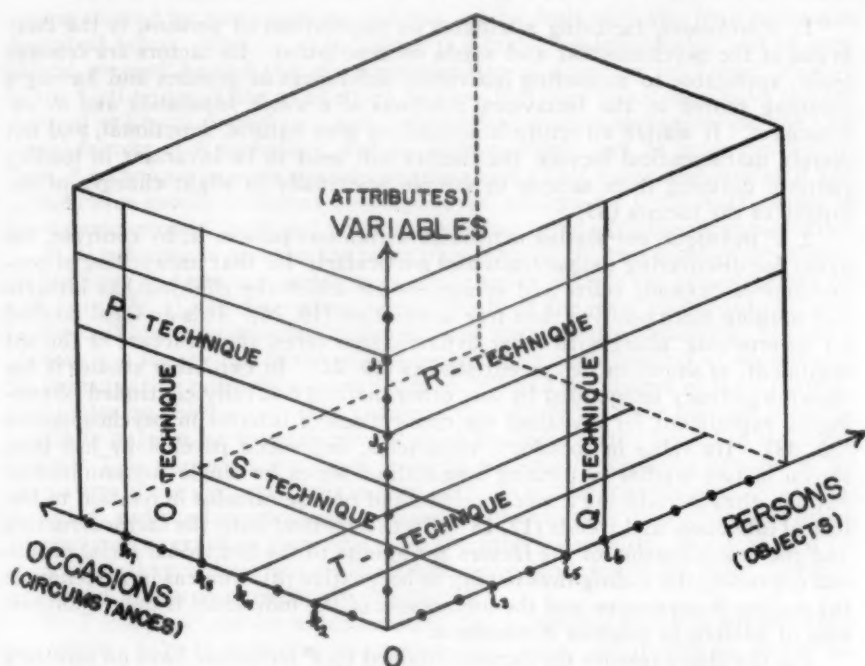


FIG. 1. THE COVARIATION CHART

series belonging to a single class. Thus in addition to the correlation of tests j_6 and j_7 for a series of people, as just illustrated for the classical R technique, we can draw a channel to the left (Fig. 1) which represents a correlation of j_6 and j_7 upon a series of occasions $k_0 \cdots k_n$, for one person i_0 .

Or again, we can take two occasions, as shown at k_2 and k_4 and correlate the series of people $i_0 \cdots i_n$ on a test j_0 , as labelled T technique. This, incidentally, is a reliability coefficient, and a whole matrix of such pairs of occasions could be factorized to find "factors in occasions (circumstances)" producing similar behavior on a test. Channels drawn in any one *plane* amount to correlatable series in which the same thing is held constant. For the present we propose to refer only to rectangular series drawn parallel to an edge, omitting the special problems of "staggered" (lead and lag) correlations, etc.

THE UTILITIES OF THE BASIC DESIGNS

It may help to fix the above six designs in mind, for the purpose of further abstract reasoning about them, if we expand briefly on the special scientific utilities of each.

1. *R technique*, factoring attributes on populations of persons, is the daily bread of the psychometrist and needs no description. Its factors are *common traits*, applicable to measuring individual differences of persons and having a meaning rooted in the behavioral relations of a *whole population and its environment*. If simple structure is applied, to give natural, functional, and not merely mathematical factors, the factors will tend to be invariant in loading pattern, differing from sample to sample principally in slight changes of obliquity of the factors (35).

2. *P technique*, correlating attributes within one person, is, by contrast, the agent for discovering *unique traits* and particularly for that unravelling of connections of dynamic traits and symptoms for which the clinician has hitherto had nothing more positive than free association (19, 20). It is an ideal method for determining, along with other dynamic structures, the structure of the self sentiment, as shown in two recent studies (19, 22). In two other studies it has shown a potency unequalled by any other method but fully controlled physiological experiment for revealing the connections of interest in psychosomatics (20, 38). Its value in sociology, economics, and social psychology has been shown in two studies factorizing longitudinal series for single communities or nations, thus introducing a more positive and precise calculus in relation to historical influences and trends (17, 18). By this method both the factor structure and the quantification of the factors are unique to the individual social or animal organism, the scaling thus having to be ipsative (8), whereas in *R technique* the scaling is normative and the uniqueness of the individual is only a uniqueness of pattern in *common* dimensions.

For the above reasons the factors obtained by *P technique* have no *necessary* mathematical relation to those obtained by *R technique*. The half dozen pioneer studies so far published suggest, however, that a *scientific* relation will exist in that the structural patterns of the individual unique factors will tend to scatter about the loading pattern of the corresponding *R technique* common factor. The *R* and *P* factors may also differ in total variance contribution. For example the surgency-desurgency factor of elation-depression seems to stand higher in factors of intra-individual variance than in inter-individual variance (13, 19, 20, 38).

3. *Q technique* has its chief use as a classificatory device for finding the subpopulations in a nonhomogeneous population (like Lazarsfeld's "latent structure analysis" or latent subgroup analysis), and for the special purpose of quantifying the extent to which an individual may be regarded as belonging to certain species and subspecies. Whether this can be considered a process of defining types depends upon the meaning one assigns to the term "type."¹ In social psychology it has value in picking out roles, i.e., common patterns of social re-

¹ I have suggested elsewhere (8) that "type" is being used for four distinguishable concepts: (a) continuous types, e.g., a tall and a short type of man, where a whole factor pattern varies in level of elements continuously from extreme to extreme; (b) discontinuous types, e.g., man and pigmy, where the pattern is essentially the same as in continuous types but the measurements are bimodal or discontinuous; (c) continuous species types, where the pattern itself differs, but in a continuous way, e.g., business men and artists; and (d) discontinuous species types, e.g., dogs and ducks, where the pattern is distinct and discontinuous as to distribution in nature. Some uses of *Q technique* to find types slur these differences.

sponse showed by many persons; but this can be dealt with also by *S* technique described below. Strictly, *Q* technique is a factor analytic design and as such should proceed *beyond* the mere examination of a matrix for "type" clusters to the further derivation of abstract factors. If it does this it may have the additional utility of providing a new, independent avenue to discovering the universal factors of *R* technique, and an avenue which may be more convenient than *R* technique in some special circumstances. This transposability of *R* and *Q* technique results is denied by Stephenson, though Sir Cyril Burt claims to have demonstrated it (6) and some statisticians agree with this demonstration.

4. *O* technique is perhaps the more important of the two *occasion-correlating* techniques. It tests one person on a whole series of tests for two occasions (k_6 and k_7 in Fig. 1) and determines the similarity of the total person (or, at least, as much of him as is included in the series of tests) on two occasions. The most obvious use for the factorization of a matrix from such pairings of occasions is for investigating multiple personality (9, p. 99) and the change of the self-structure under psychotherapy. The nature of the factors has to be inferred from the attribute pattern for those occasions which are most highly loaded in the given factor. This approach also lends itself to analysis of stimulus situations, for the nature of the circumstances in each occasion can be manipulated, or at least recorded, and the factors among occasions will thus show how situations group themselves in regard to their effects on the total personality. Some alleged instances of *Q* technique are really *O* technique, and would be more correctly evaluated if this were explicitly recognized.

5. *T* technique may be remembered by the mnemonic that it is *test-retest* factorization. Like *O* technique it correlates occasions, but it does so, as the covariation chart above shows, by repeating the same test for a matrix of different occasions on a population of persons. It is thus a factorization of reliability coefficients, but naturally the significance of the design would hinge upon recording the particular *circumstances* of each occasion of re-administration, as in *O* technique. The factors among such defined occasions would again be factors in the general stimulus situation as judged by effects on the population. But they would differ from *O* technique in being factors with respect to some single response, e.g., an opinion poll on a definite issue, instead of to a sample of the responses of the total personality, and in applying to a population instead of a single person. Because of this analysis of circumstances according to the effects they produce on the distribution of response with a population, *T* technique might be called a "social climate" thermometer and is likely to find its most valuable application in quantifying phases of business cycles and in getting more historical meaning out of opinion polls than the present poverty of statistical methods permits.

6. *S* technique may be remembered by the mnemonic that its principal immediate use is to detect and define *social roles*. It correlates two persons on their reactions to a single stimulus (test) on a series of occasions. (A role is defined as a pattern of responses to different occasions which is modal among individual patterns, i.e., it is a cluster or factor among people in responses to social occasions.) The loading of a person in the factor would show the extent to which he is successfully assuming the role. *S* technique also has promise in determining the internal structure of a group or institution; for example, by correlating the responses of members of a family over many occasions on their response to a single issue it would indicate the functional subgroups within the family. Be-

cause of this capacity to reveal the degree to which individuals belong to groups characterized by homogeneity of functioning in regard to a series of historical events, *S* technique is *par excellence* the method for social psychology.

THE INTERRELATIONS OF THE SIX TECHNIQUES

In respect to inherent, systematic characteristics there are in fact two main kinds of relationships to be considered among the six techniques, and these are again most readily perceived by reference to the covariation chart itself (Fig. 1).

First, we observe that two techniques may depend on correlations which lie in the same face of the parallelepiped, which means that their correlation matrices share the same entries, i.e., they work upon the very same data but correlate it in different arrangements. This gives us the three pairs of groupings *R-Q*, *T-S* and *P-O*. Thus the first pair, *R* and *Q*, share a matrix of test measures upon people; *P* and *O* share a matrix of test measures upon occasions, and so on. (Or alternatively, they hold constant the occasion, the person, etc.)

Second, any two techniques may share an edge of the model, i.e., the parallels representing their correlations hinge on the same edge. This means that they are seeking relations among the same variables but upon different modalities of population. The correlations of the same variables in the two techniques have different meaning because they deal with different influences and a different source of variation, yet because they are the same variables both techniques are required as supplementary statements of the nature of the factors in the variables. This groups the six techniques in the three pairs: *R-P*, *Q-S* and *O-T*.

Now the first of the above groupings distinguishes *among* the pairs on the grounds of modality of data, i.e., the members of a pair work on common referents, while *within* the pairs the distinction is one between complementary statistical processes. Indeed, we shall argue below that since the complementary processes rearrange the same data their results must in general be statistically transposable and that the members of a pair are really equivalent. The researcher's interest in bringing the two methods of one pair to bear upon any problem is therefore to obtain statistical confirmation of the same scientific relations, and to check the soundness of the statistical part of his research by proceeding from data to conclusions through two routes.

By contrast, the pairings according to the second principle guide him to possibilities of independent experimental or general scientific confirmations. Thus *R* and *P* techniques have been used with precisely the same batteries of personality measures (9, 12, 19, 20) but, respectively, upon a whole population at once and upon a series of individual

studies of clinical cases. It has been found that some of the dynamic unities revealed by the one are also revealed by the other (11, 14, 17, 22). Similarly in the *S-Q*, and *O-T* pairs of techniques we may in general expect that connections between variables in one population will be borne out if the functional unity of the trait is universal by their connections also in other realms of variation.

For clarity of further reference we shall call the three pairings of the first type *common matrix techniques* (yielding *transposable factorizations*) and those of the second *common variable techniques* (yielding independent but *common function factorizations*). Thus *R* and *Q* are common matrix techniques because both begin, or can begin, with the same score matrix.² In the first case one correlates the columns (tests) and in the second the rows (people). It is the writer's opinion that psychological research today stands most in need of appreciation of the gains possible from a two-handed use of the *common variable techniques* on any one scientific problem in mutually checking studies. But since some current bandwagons of enthusiasm have stressed the common matrix designs to excess, it is necessary to digress into considerations which will help perspective.

The misperceptions that, in the writer's opinion, are most prevalent are (a) those shown in attempts to obtain by *Q* technique what is better obtainable by *P*, *O* and even *R* technique, and (b) in a denial of any statistical relationship between the results within each pair of common matrix extractions, e.g., of *R* and *Q* techniques. As to the latter we should note that the original development of *Q* technique (though not under that label) by Sir Cyril Burt in 1912 (2), 1917 (3), 1931 (4), and 1933 (7) occurred expressly with the intention of arriving at the ability factors simultaneously hypothesized in *R*-technique studies. In Stephenson's 1935 article (29), "Correlating Persons Instead of Tests,"

² It seems to have been misunderstood by Stephenson that this label refers to common form of matrix, not necessarily to such furnishing as hypotheses, or the subjectivity or objectivity of data, number of entries, distribution of variables, etc. In designing a test-person matrix for *R* technique the number of persons, *N*, will usually be larger than the number of variables, *r*, while for *Q* technique the number of available persons, *n*, is usually small and the range of variables, *R*, is large. Naturally this difference of emphasis will also be associated with differences in sampling of variables and persons and the choice of hypotheses. Such differences for that matter will also occur within examples of *R* technique itself, according to the particular design and purpose of the experiment. But the *essential* fact is that both use a test-person matrix, and if one wishes, as in a recent study by the author using 100 tests on 100 persons, the *identical* matrix can be used, first read upright and then sideways. Stephenson's quibble would amount to saying that since some books are taller than they are wide and other the converse, it is a great mistake to recognize that they are all books!

Burt's design was acclaimed with new enthusiasm but lost some of these earlier perspectives and insights. Incidentally, Burt (5) credits the germinal idea to Stern (32) who in 1911 wrote of "horizontal correlation" and "vertical correlation," though without the superstructure of factor analysis, and he points out also an early use of Q technique by Beebe-Center in 1933 (1). Incidentally, Stephenson titled his basic article "The Inverted Factor Technique" (30), i.e., the inverse or obverse of R technique. However, statistically-minded psychologists have preferred *transposed* factor analysis, for, to be algebraically exact, Q technique factorizes the *transpose* of the R -technique matrix.

Accordingly, here and elsewhere the writer follows the convention of calling Q transposed R technique, while S is similarly transposed T technique and O is transposed P technique. If we are correct in supposing that the transposition is a purely statistical operation—a matter discussed in the next section—there are basically only three independent factor-analytic experimental designs, namely, R , T , and P techniques. These alone yield factors in statistically independent, unbridgeable measurement systems, and are the pillars between which bridges of paired scientific inference and relation can be built.

THE MEANINGS AND UTILITIES OF THE COMMON MATRIX TRANSPOSES

It is necessary to stress in the penultimate sentence above that we speak of experimental designs not experiments. For an R - and a Q -technique experiment using samples of tests and persons from the same populations would be independent experiments, but only in the sense that two R -technique experiments on related data would be independent.

The belief of some users of Q technique that it is fundamentally different from its transpose technique— R —and, indeed a method *sui generis*, has so far been most exhaustively statistically examined and refuted by Sir Cyril Burt (5). Saunders' work on "direct factorization of score matrices" (27) supplies an incidental proof of their interdependence from a new angle. In the writer's experience professional statisticians take the position that there is no doubt about the transposability of factors from a double-centred score matrix though there may be doubt about the exact relation under other and special conditions. A paper by Hicks³ and the present writer aims to deal with these statistical issues in more detail.

³ HICKS, V. E., & CATTELL, R. B. The problems of transposing factor solutions in common matrix factorizations, illustrated by an example in R and Q techniques. In preparation.

Here it is possible to examine the relations of transposes only in an introductory and illustrative fashion, as far as method is concerned, proceeding mainly from first principles in the use of the correlation coefficient. However, we shall do so comprehensively from the standpoint of experimental design, under five headings as follows.

1. *Transposability of variance findings, with partial loss of variance information.*⁴ It is relatively easy to see that R and Q (or P and O , T and S) techniques normally (i.e., without double centering) have the completeness of their transposability slightly restricted by some inevitable mutual losses of information. The losses which then occur are (a) of the variance of the first factor (or in some conditions the first two) and (b) of the specific factors.

The first kind of loss can be most readily perceived by a concrete example, say that of correlations upon a matrix of physical measurements, e.g., leg length, size of shoe, sitting height, weight, for a population of men. If we correlate such "tests" the first R -technique centroid factor is likely (and indeed known) to be a "general body size factor," since the man with greatest stature is likely to have the largest boots and the greatest weight. But if we correlate persons, nothing corresponding to this is discovered, for a small man and a large man of similar proportions will correlate perfectly. Size is overlooked, because the correlation coefficient "responds" to similar profiles, regardless of level.⁵ The correlation coefficient in effect behaves as if it has scaled the raw scores to standard scores, as far as the columns being correlated are concerned, reducing them to the same mean and the same sigma.

Reciprocally, R technique loses a factor through virtually bringing the tests to the same means and sigmas. The fact that all men are taller than they are wide is overlooked by the correlation of height and width, but not in the corresponding (raw score) Q -technique procedure. The first centroid in Q technique is largely due to the correlations arising from the fact that all people resemble each other to some extent, e.g., they are all taller than they are wide. This massive factor in the matrix may be called the "common species" factor, for it defines the common pattern of the population and the individuals's endowment therein represents the extent to which he resembles the species.

The relative importance of these reciprocal losses will be argued below, but let us at least notice here that most psychologists, on common-

⁴ This section is concerned with the difficulties referred to in Stephenson's comment that "it has long been difficult for everyone except myself to accept the proposition that R and Q are obviously completely different." (Contribution to a symposium held at the meeting of the Midwestern Psychological Ass., Chicago, Ill., April 27, 1951.)

⁵ The r_p "coefficient of pattern similarity" (18) takes account of difference of level as well as of shape. If we had some experience with its use in matrices for factor analysis it might transpire that an improvement of Q technique would be obtainable through substituting r_p for r .

sense grounds, are inclined in *Q* technique to *standardize* the rows of tests (they are rows in the transposed matrix) before correlating the columns, for they point out that the raw score in which a test is couched is arbitrary and irrelevant. For example, if, in the above illustration, body weight is measured in ounces, this variable will be highest for every individual, whereas length of nose (in feet) will be lowest for everyone. The important thing, it will be argued, is whether the person's nose is large for a population of noses.

Such a standardization of rows when one is correlating columns of persons, or, in *R* technique, a standardization of rows of labelled persons when one is correlating tests, takes the *species factor* out of *Q* technique and the *size factor* out of *R* technique, so that both now lack the same two "potential" factors and, for the rest, have factor solutions which could be truly transposable. The general statement that it is possible to use either technique as a gateway to the same end result, therefore, needs the modifying clause that with usual scaling procedures one (or two) general factors are lost in the process.

The relative importance of the factors reciprocally lost in the transposes remains to be evaluated. For most purposes the "species" factor lost in *R* technique in the present example is less important than the "size" factor lost in *Q* technique, for we know the species with which we are dealing and are most interested in individual difference within the species, having nothing to do with species character. Furthermore, though *Q* technique necessarily loses a second factor, for reasons given below, there is really no need to lop off the second factor in *R* technique, so that the latter in fact loses no information of interest and value in the predictions with which psychology is chiefly concerned.

The first potential *Q*-technique factor—that of general size in our present example—is lost inevitably. The second is pruned away by that use of standard scores which most psychometrists rightly demand in this situation, wherein the absolute scores of variegated tests are meaningless. But in *R* technique the corresponding pruning—by standardizing persons instead of tests—is not only unnecessary but objectionable. Bringing different tests to the same mean and standard deviation, despite wide differences in raw scores, makes sense because there is no meaning to such a statement as "Mankind as a whole is higher in mechanical aptitude than intelligence."

Let us next briefly examine the situation regarding specific factors, i.e., those peculiar in *R* to a test and in *Q* to a person. In the first case a group of people covary in their performance on one test only, and in the second one person has a covariation on several tests not shown by anyone else. It would seem that these are inevitably lost in the transformation. Simultaneous variation on tests and persons must exist for transposition to occur. Light is thrown on this by Saunders' *K*-way scale analysis (27), proceeding directly from the score matrix.

The mutual losses in *R*- and *Q*-technique transformations should

be further considered by the specialized reader in the general perspective regarding losses of information when factoring (a) covariances, (b) score matrices, and (c) correlations, or, indeed, in the perspective of statistical abstracting generally.

2. *Transposes examined in terms of experimental convenience.* So far the discussion has proceeded as if we dealt with a roughly square matrix, correlated on a set of columns for *R* technique and turned on its side and correlated on what were the rows for *Q* technique. But the labor of dealing with as many variables as one needs to have people, for a reliable sampling of the population, is too great and, in practice, the score matrix has always been trimmed to oblong form. Therein the short side—that representing the things to be correlated—consists of tests in *R* technique and people in *Q* technique. The resulting failure of symmetry adds further differences between the two techniques though they are differences of degree, not of kind. What we need to discuss under this heading concerns the claim by some clinicians that the same result may be obtained with less labor by *Q* technique. Actually, if we preserve statistical equivalence (equivalence of reliability), the decrease in subjects is balanced by an increase in tests and the number of subject-hours of testing time (or the cost of subjects) remains exactly the same. Since much testing can be group testing, however, there is more saving of the experimenter's time through *R* technique.

The respective availability of time and subjects, therefore, may well dictate the choice as far as this very practical consideration is involved. But there are also questions of "convenience" from the standpoint of reconnoitering hypotheses and the strategy of research. These will be discussed below in relation to analysis of variance and testing of hypotheses.

3. *Transposes in terms of generalizability of findings.* It will be evident that despite the complete *logical* symmetry existing between the transposes we have to admit at several points qualifying conditions on the reciprocity, owing to different inherent properties in the "equivalent" series, etc. This needs to be kept in mind most steadfastly in regard to the sampling problems and the resulting restrictions on generalization. For the sampling of tests, with our existing knowledge and concepts, cannot be handled in the same way as the sampling of *people*.

The argument that restricted sampling of people is more dangerous than restricted sampling of tests does not seem to have been seriously appreciated everywhere, though it has been duly regarded in Burt's original writings on *Q* technique. Stephenson's position⁶ appears to be

⁶ His assertion that even a single case (a crucial experiment) can prove or disprove a general scientific law seems to the writer to confuse the distinction in scientific method between controlled experiment, reaching its extreme expression in some experiments in physics, where all irrelevant variance is controlled, and statistical analysis of variance *in situ* and uncontrolled, as it has to operate in most social science research (16). It may also confuse description and explanation.

that if he can establish certain patterns of relation among people in a small group he need not be concerned with their wider generalizability. This betrays a fundamentally different assumption about the aims of science, substituting a descriptive and revelatory (one might almost say artistic) goal for the usual one of explanation and generalized prediction.

If only half a dozen persons are taken in *Q* technique the results can be generalized to the parent population only with the high degree of guesswork which that very small sample permits. The fact that tests have been multiplied does not save one from the laws of sampling inference where people are concerned. It should be clear, however, that this argument does not deny that the same general life processes and logic actually operate in a small group, or a single person, as operate in the parent group. It only denies that such methods of investigating small groups or individuals, without regard to statistical needs, *are capable of finding them*. Indeed, the boot is on the other foot, for it is the present misusers of *Q* technique who deny that *universal* processes, particularized to the situation, explain the behavior in the small group or individual; for they claim that the processes which *can* explain are understandable without any experience of their manifestations elsewhere. To which one can only reply with the logician that the meaning of "A" requires knowledge of "not-A," or with the poet "What know they of England, who only England know?"

Naturally we should ask if an equivalent objection applies to the reciprocal situation of the small number of tests used in *R* technique. This is where statistical symmetry is no longer a scientific or epistemological symmetry. *The test population, except for the personality sphere concept (9), does not have the qualities of a biological species population.* The experimenter, indeed, usually does not even make any attempt to sample the test universe. He establishes factors in a given area, staked out by defined and preserved tests, and then, in later experiments, he goes on to extend the area. He does not say "These are all the personality factors in existence," but "These are the factors *only in the area I have so far staked out* with the defined tests, *but* they are true of *all* people." The equivalent procedures for *Q* technique are by no means so dependable, e.g., one cannot preserve people as "markers," as one does specific tests.

There is thus in any pair of transposed common matrix techniques a tendency for one of them to rest on a surer sampling foundation than the other and to permit wider generalization with the same amount of work.

4. *Comparison of transposes in terms of reproduction and interpretation of results.* As indicated incidentally above, the interpretation of factors and the preservation and communication of defining data are undoubtedly rather less efficient in *Q* than *R* technique. Where *R*-

technique factors require the filing of two or three tests and a note on the population sample, *Q* technique requires that we file a very lengthy profile of performances for the highly loaded individuals—or file the individual himself!

Interpretation of the factor in *R* technique requires that we contrast the nature of the highly- and the zero-loaded test. This is at times a fairly difficult process of abstraction and inference, but it does not present the double act of abstraction required by *Q* technique, where we compare a whole series of test scores for the person of high loading with another series for the individual of low loading.

To examine the living person himself is no easy alternative. For although a person may be, say, highly loaded in the factor of intelligence, his every act presents intelligence inextricably interwoven with every other ability, emotional, and temperamental factor in his nature; whence recognition of what "characterizes" him is not easy.

5. *Simple structure and factor invariance in the transposes.* Assuming that invariant, stable factors have so far been producible only by rotation for simple structure, the problem of how effectively invariance can be obtained in the transposes reduces to an examination of what simple structure means in each. The scientific assumption that most factors will operate significantly only in a limited number of any widely chosen battery of test performances, on which *R* technique rotation rests, is logically acceptable and empirically verified. In *Q* technique it is logically less acceptable that some *persons* will be *wholly* devoid of a factor operating in others.

Discussion of rotation requires the statement of our general position on *Q* technique, which is that its true contribution (apart from occasional convenience in approaching *R*-technique factors through a transpose) is specifically with *nonhomogeneous populations*. Then *Q* technique or the method of latent (structure) subgroup analysis, carried perhaps only to the point of finding correlation clusters, can be used to find the separate "species types" or homogeneous subgroups upon which *R* technique can be most profitably employed. Thus in a group composed of, say, trained artists, soldiers, and philosophy students, a simple structure rotation could be expected, *provided the variables deal only with those aspects of personality which are much affected by occupation*.

Apart from such nonhomogeneity one would expect the distribution of *Q* factor loadings to have a substantial proportion of zeros only because of the effects of normal distribution of loadings.⁷ Unfortunately

⁷ The frequency of essentially zero loadings in a well-sampled *R*-technique battery is typically around 60–70 per cent with respect to any one factor (10, 12, 14), which should be distinguishable from the smaller percentage of apparently zero loadings resulting from the application of a smooth normal distribution curve to the given factor loading distribution.

there seem to be no *Q*-technique studies substantial enough to have been examined for empirical evidence as to an invariant rotation position.

If no universally acceptable method of ensuring invariance in *Q*- and *S*-technique factorizations should be found, their whole utility falls to the ground and they cannot be used as alternative avenues to the general universe of scientifically negotiable results provided by their transposes, *R* and *T* techniques.

DERIVATIVE FORMS RELATED TO SCALING AND SUBJECTIVITY-OBJECTIVITY OF EVIDENCE

The discussion has so far dealt with the patterns of correlation possible by series arrangeable among the three chief signature characters of a measurement, but it is evident that, if the two remaining signatures are also considered, fairly numerous additional combinations are possible. For practical purposes only two or three possible categories actually make sense in these series, and the total of six designs so far considered is in the end only doubled or quadrupled. However, it is most important to deal with these further modifications because in some recent writings they have been confused with the primary differences, e.g., *Q* technique has been said to differ in terms of scaling (31).

Scaling differences rarely affect factor-analytic conclusions in themselves. But let us consider the principal alternatives, which are: raw scores, normative scores (ipsative (8) within a single person, as in *P* technique), scales of the Guttman type, logarithmic scales, percentiles and raw scores forced to a normal distribution. The only modifications relevant to factor analysis are the forcing of arbitrary raw scores to a normal distribution, which Thurstone (36) has shown to give clearer factor structure, and the change to normative scores across the *arrays* (rows or columns) correlated, which we have shown above to eliminate a factor. Stephenson (31) speaks of four "foundations of psychometry," meaning standardizing rows, standardizing columns first and then rows, and so on. As indicated, the two last are superfluous, for the use of *r* automatically produces standardization one way. The real alternatives are not to standardize at all or to standardize across the arrays being correlated. Much discussion at cross purposes might be saved by a convention of writing these as *Q* and *Q*₁ or something similar.

The distinctions connected with the fifth signature of a measurement (nature of the observer) are, however, more fundamental than those of the fourth (scaling). Some workers in personality factorization have adopted the convention of writing *BR*, *Q*, or *OT* alongside the factors (9, 10, 12, 13, 21, 22, 35, 38) to indicate respectively whether they rest

on behavior rating *in situ*, on questionnaire data, or on objective (non-introspective) behavior data in the special situation of a test. From numerous scientific standpoints the results from these three sources need to be differently regarded and differently treated, but from the broader viewpoint of the basic character of the observations the difference between behavior *in situ* and behavior in a specially devised test is irrelevant and we have only the distinction between behavioral data and subjective, introspective, questionnaire data (the latter when treated at face value).

Broadly speaking the present writer is in agreement with such behaviorists as Spence (28) that introspective data cannot be truly integrated into scientific psychology. The extensive use made of such "data" in some recent *Q*-technique work is consequently abortive. But this evaluation must be qualified by the recognition that (a) the "mental interiors" of factors (12, 14, 21) are of interest even if we cannot integrate them and they remain mere epiphenomena, (b) in reconnoitering stages of research and with subjects of blameless motivation they are a rough guide for later objective research, and (c) the objection disappears if the introspective reports are treated only as behavior. For the scientific restriction is not on using the results of introspection, or on questionnaires, or on verbal responses, but on taking as data that cannot be witnessed by any second observer. In other words, in factor analysis this amounts to using correlations where we cannot get a reliability coefficient calculated between data obtained by two distinct observers.⁸ *Verbal responses are acceptable if the experimenter refrains from accepting the conventional face-value of the words as symbols and independently establishes their relation to other behavior, which could best be guaranteed by getting them in Urdu or Swaheli or some other language he does not understand.*

In the more systematic *R*-technique studies the meaning of questionnaire (or *R*.) factors has been established by their correlation with behavioral factors (21), but in recent *Q*-technique studies of the self-concept the temptation has prevailed to accept introspection instead of inferring the self-evaluation from behavior. Since, as Freud and sundry philosophers have observed, language is not only a limited but also a systematically distorted form of behavior, a correlation expressing a person's resemblance to another (*Q* technique) or a person's resemblance to himself on another occasion (*O* technique)—even when language is treated strictly as behavior—has all the unreliability and meaningless-

⁸ The observers must be peers in objectivity. If even one is introspecting, the whole belongs to factorization of the subjective.

ness of working with a quite inadequate sampling of the relevant parent "population of variables." Actually the personality studies of Stephenson (29) and Rogers (26), as argued in a recent article to clinicians (15), are Q_r - and O_r -technique studies, not integratable with behavioral factor analysis. For though statistical treatments of introspection are interesting, as agreed above, they relate to behavior only as a non-Euclidean to a Euclidean geometry.

Burt has pointed out that the general negotiability of Stephenson's Q_r technique is additionally confused by the almost mystical instruction on "significance" which every subject, bright or dull, is required to understand in order to participate in the experiment. When subjects are asked to rate traits according to their "significance" to their personalities, one may surely expect that the differing subjective perceptions of this instruction will add an additional dimension of error to the subjectivity already inherent in the replies. R_r technique at least avoids this difficulty. The best statistical meaning that can be given to "significance" is that the subject is being asked to rate the various aspects of his behavior or feelings on a continuum of "eccentricity" or deviation of those kinds of his behavior relative to other kinds—all in relation to an implied norm. In R_r one asks him the difficult question of how deviant he is in say "liking Rembrandt" or in "computational skill" relative to the general population, but in Q_r we pile on a demand for the further judgment of how much more deviant he is in liking Rembrandt than in his computational skill!

THE CONTEXT OF FACTOR TECHNIQUES IN GENERAL SCIENTIFIC METHOD

This concluding section proposes briefly the setting of current factor analytic designs in statistical and scientific method.

As stated at the opening of this article, the observations of covariation which suggest or test scientific hypotheses may be either (a) univariate or (b) multivariate. In the former the variable observed is the *dependent variable* changing in response to controlled changes in the independent variable. In the latter the variations of all are observed simultaneously with no temporal or other priority for any one of them. The former has usually been associated with actual experimental manipulation whereas the latter is of value in psychology and sociology, where some of the most important things cannot be controlled in the laboratory but have to be examined *in situ*. The former *tries* to keep everything constant but the pair of variables concerned, and what it cannot control it calls *error*. The latter controls little or nothing but keeps

track of what actually happens to many variables which would be unknown sources of error in the former. (It still leaves some unknown variance in specific factors.)

The statistics of the former are mainly concerned with significances of differences of means and analysis of variance, while the latter uses multiple correlation, the discriminant function and, principally, factor analysis. For brevity we may discuss factor analysis and analysis of variance as representatives of the two halves of statistical method which need to be related if one is to see more clearly how factor analysis functions. Incidentally, the present writer (16) has proposed a hybrid of these two in which some degree of experimental manipulation and control of variance is combined with factor analysis.

The statistical differences of factor analysis and analysis of variance are embedded in profound differences of strategy in scientific method. In the latter the experimenter assumes he can test (or produce) his hypothesis by a single variable (or pair), i.e., that he knows already which variable is most important. In the former he realizes that in a bewildering array of variables the significant variables (factors, latent variables, variables of greatest influence in the field) have still to be found. What is more, he recognizes that frequently the concept in a hypothesis cannot be operationally defined in a unique way by one variable. It is a paradox of this methodological contrast that whereas in analysis of variance the experimenter generally assumes more insight into the relevant hypotheses, he applies a less severe test to them (trusting to a one-variable manifestation instead of a pattern) and gets the information only that "a significant relationship exists" instead of a quantitative statement of the degree of relationship, as in correlation and factor analysis.

Looking at psychological research of the past thirty years one cannot avoid the conclusion that factor analysis should more frequently have preceded controlled experiment and analysis of variance. For example, there have been countless experiments in learning relating dependent variables to degree of hunger, and in personality in relating life adjustments, etc., to degree of extraversion. The experimenter has generally been content to represent each by one variable (hunger by hours of deprivation, extraversion by some arbitrary rating from Jung). A factor analysis should first have been performed to see if a factor pattern could be found corresponding to each of these concepts and to find the pattern of weighted variables most accurately estimating each in further experimental work.

Since the object of juxtaposing this article with Stephenson's in the

present issue is to bring out differences of viewpoint, I must turn to what seem to me misrepresentations on his part of the above discussed roles of factor analysis in relation to analyses of variance and scientific method generally. Stephenson's proposal to relate analyses of variance to *Q* technique does not seem to me to add anything not already recognized in these methods. It does not present the novelty of experimental design found in the more radical hybrid proposed above (16). As an expanded and more brilliant treatment of the sketch just presented of the relationship of the two methods, the reader is referred to a recent article by Burt (6) who points out that analysis of variance may be used to demonstrate but not discover factors,⁹ to test their significance but not their nature (pattern).

But the most misleading assumption is that *Q* technique has a monopoly (presumably among the six factor-analytic designs) of "hypothetico-deductive direction . . . singular propositions . . . transitory postulates"—in short of hypothesis testing! Stephenson's assertion that hypothesis testing "can find [no] place in *R* technique"¹⁰ can be refuted without going further afield than the work of my immediate colleagues, where the design and choice of variables in objective personality tests (10) and the investigation of ego defense mechanisms (37) were guided by several highly explicit, testable hypotheses. Or again, in the first experiment (19) relating *R* and *P* techniques it was hypothesized that the same factor structure would be found in any one person as in the general population, and every single variable and condition was carefully chosen in relation to this theory and five specific subhypotheses. Or, yet again, we may note Stephenson's still more naive assertion that *Q* technique enables us "for the first time in history to operate with testable assumptions about the self."¹⁰ It has evidently escaped his notice that *R*-technique studies, with objectively measured attitude-interest variables have already (13, 14) been designed on a set of explicit hypotheses about the structure of the self-sentiment and have emerged with contingent proof of the most important of these.

All variants on the three basic factor-analytic designs—*P*, *R* and *T* techniques—share the power of factor analysis to *produce* hypotheses more readily than univariate methods and to *test* them more searchingly. The regularity of covariance, the lawful behavior, observed in a

⁹ A group of tests (or a set of persons) sharing high endowment in the same factor will have (standard) scores with significantly less within-group variance than between-group variance with relation to other groups also largely made of one-factor measures.

¹⁰ Quoted from Stephenson's paper which was read at a symposium held at the meeting of the Midwestern Psychological Ass., Chicago, Ill., April 27, 1951.

pair of variables cannot suggest such overdetermined hypotheses as can the emergence of a factor.¹¹ Re-entering a factor-analytic experiment with the hypotheses corresponding to the factors first found permits a more searching test of them, because if a given hypothesis is to be verified, not one variable but a whole pattern of variables has to behave in the manner predicted by it.

It may be questioned, however, whether *Q* technique as used by Stephenson and Rogers (*Q* sort) is getting the full advantage from the hypothesis-producing power of factor analysis. When Rogers starts with such a hypothesis as "where the self-concept is formed entirely from the evaluation of others the individual will at some point face internal conflict" (26), one is evidently dealing with something resulting from a casual glance of the naked eye rather than a precision concept gained by the factor-analytic microscope.¹² And when Stephenson proposes to "take the guesswork out of factor analysis" by using only variables from a preconceived theory—gained from an inward eye, or Jung, or Aristotle—he is neglecting the more precise hypothesis formation possible from first seeing what lawful relations factor analysis will show in a truly varied and comprehensive array of variables. To most American psychologists this Old World subjectivity amounts to putting guesswork *into* factor analysis, not taking it out. In too many of these discussions we are dealing with an inverted semantics, not merely an inverted factor technique.

SUMMARY

1. There are six primary factor-analytic experimental designs, defined as *O*, *P*, *Q*, *R*, *S* and *T* techniques.

2. As far as statistical independence is concerned, these reduce to three independent *common matrix* pairs: *R(Q)* technique, *P(O)* technique, and *T(S)* technique. In each pair, one is the transposed technique of the other, and, with attention to required conditions, can be

¹¹ This is the logical point at which to make reference to the fatuous, but, one hopes, now less fashionable, slogan that "one only gets out of a factor analytic design what one puts into it." This is actually a truism of all experiment—one cannot find relations between variables one has not used in the experiment. But the hypothesis about structure with which one emerges may be entirely different from those with which one entered the experiment, e.g., one may find seven factors where it was asserted there were only two.

¹² How many people exist whose self-concept is formed *entirely* from others? And is it not safe to say that *everyone* faces internal conflict? One would want to know what factorial proof has been given of the unitary dynamic nature of this alleged entity the "self concept." And, of course, the *degrees* of conflict associated with the *extent* of its formation by associates could be most readily investigated by *R* technique.

used as a second avenue to the discovery of the same factors (though a fraction of the group and all the specific factor information are lost in the process).

3. The six experimental designs can also alternatively be reduced to three *common variable* (and therefore common loading pattern) pairs. Within these there is no *statistical* transpose equivalence as in common matrix pairs, but a scientific *experimental* equivalence in that the same functional entities may be tested in two different contexts of manifestation.

4. Since any measurement has five basic referents or signatures which particularize it, the above six combinations among three of the referents (forming the covariation chart) can be multiplied according to further reference to (a) scaling or (b) subjectivity or objectivity of data. Scaling possibilities extend from the use of covariances instead of correlation to the use of doubly standardized score matrices, but these differences have no *major* influence on the results. Subjectivity of data, i.e., the impossibility of reliability coefficients among an extended series of different observers, operates, however, to remove Q , O , etc., results from negotiability in the universe of behavioristic psychology.

If two varieties of scaling and two of data are considered along with the major variants, this creates 24 distinct factor-analytic realms of factors, which need to be distinguished by a conventional symbol system since they are never precisely equivalent and the failure to distinguish them—which is now prevalent—causes considerable confusion.

5. The choice among the three basic (P , R , T) techniques is best directed by the realm of phenomena to be investigated, while choice between the alternative transposes in each is determined by convenience, e.g., population and time available; by purpose, e.g., interest in a large or small range of variables; and by reliability of results, e.g., as to rotation, preservability, and interpretability. As to these last, Q technique is less satisfactory than R technique.

6. Q technique has its greatest usefulness in detecting and defining species types in a definitely nonhomogeneous population.¹³ Where the population is reasonably homogeneous it is important to apply R technique first in order that the variables used in subsequently defining types shall be properly sampled from the total range of description, i.e., shall be factor measurements. So-called Q sort commonly gives erro-

¹³ However, Burt argues that to determine the characteristics which distinguish types it is "more expeditious and more accurate to proceed by the analysis of variance" (5, p. 61). One should also consider here Lazarsfeld's latent subgroup analysis and Rulon's generalized discriminant function (when some basis for initial sorting exists).

neous values regarding the degree of resemblance of two individuals because equal weight is not given to the different factors in personality.

7. The primary purpose of factor analysis is to discover or confirm hypotheses as to the nature of underlying influences or dimensions. It proceeds to provide a specification equation for estimating (predicting) specific test performances, people, and occasions from the factors. It is more productive of relatively precise hypotheses than most other statistical methods and in general provides a more searching test of (deductions from) a hypothesis within a single experiment.

8. Factor analysis and analysis of variance represent respectively the multivariate and univariate approaches in scientific method. Hybrids can be formed between them, and one can with difficulty be made to perform some of the functions of the other, but essentially they are supplementary and appropriate for different phases of work and different kinds of research strategy. A retrospect on research suggests, however, that factor analysis might with advantage have been used more frequently to discover more significant measures of factors for inclusion in analysis of variance studies.

REFERENCES

1. BEEBE-CENTER, J. *Pleasantness and unpleasantness*. New York: Century, 1933.
2. BURT, C. L. The mental differences between the sexes. *J. exp. Pedag.* 1912, **1**, 273-284.
3. BURT, C. L. *The distribution and relations of educational abilities*. London: P. S. King, 1917.
4. BURT, C. L. *Report of Consultative Committee of Board of Education on the Primary School*. Appendix III. London: H. M. Stationery Office, 1931.
5. BURT, C. L. Correlations between persons. *Brit. J. Psychol.*, 1937, **28**, 59-96.
6. BURT, C. L. A comparison of factor analysis and analysis of variance. *Brit. J. Psychol.*, Statist. Sect. 1947, **1**, 3-26.
7. BURT, C. L., et al. *How the mind works*. London: Allen and Unwin, 1933.
8. CATTELL, R. B. Psychological measurement: Normative, ipsative, interactive. *Psychol. Rev.*, 1944, **51**, 292-303.
9. CATTELL, R. B. *The description and measurement of personality*. New York: World Book Co., 1946.
10. CATTELL, R. B. Primary personality factors in the realm of objective tests. *J. Pers.*, 1948, **16**, 459-487.
11. CATTELL, R. B. The dimensions of culture patterns by factorization of national characters. *J. abnorm. soc. Psychol.*, 1949, **44**, 443-469.
12. CATTELL, R. B. The main personality factors in questionnaire, self estimate material. *J. soc. Psychol.*, 1950, **31**, 3-38.
13. CATTELL, R. B. *Personality: A systematic theoretical and factual study*. New York: McGraw Hill, 1950.
14. CATTELL, R. B. The discovery of ergic structure in man in terms of common attitudes. *J. abnorm. soc. Psychol.*, 1950, **45**, 598-618.
15. CATTELL, R. B. On the disuse and mis-

- use of *P*, *Q*, *Qs*, *O*-techniques in clinical psychology. *J. clin. Psychol.*, 1951, 7, 203-214.
16. CATTELL, R. B. *Factor analysis: An introduction and manual for the psychologist and social scientist*. New York: Harper, 1952.
 17. CATTELL, R. B. The psychological dimensions of social progress in Britain, determined by *P*-technique, 1835-1935. *Hum. Relat.*, in press.
 18. CATTELL, R. B., & ADELSON, M. The dimensions of social change in the U.S.A. as determined by *P*-technique. *Soc. Forces*, 1951, 30, 190-201.
 19. CATTELL, R. B., CATTELL, A. K. S., & RHYMER, R. M. *P*-technique demonstrated in determining psychophysiological source traits in a normal individual. *Psychometrika*, 1947, 12, 267-288.
 20. CATTELL, R. B., & LUBORSKY, L. *P*-technique demonstrated as a new clinical method for determining personality and symptom structure. *J. gen. Psychol.*, 1950, 42, 3-24.
 21. CATTELL, R. B., & SAUNDERS, D. R. Inter-relation and matching of personality factors from behavior rating, questionnaire and objective test data. *J. soc. Psychol.*, 1950, 31, 243-260.
 22. CROSS, PATRICIA. Comparison of *P*- and *R*-technique determinations of ergic structure in common attitudes. Unpublished M.A. thesis, Univer. of Illinois, 1951.
 23. HOFSTAETTER, P. R. A factorial study of cultural patterns in the U. S. *J. Psychol.*, 1951, 32, 99-113.
 24. HOLZINGER, K. J., & HARMAN, H. H. *Factor analysis*. Chicago: Univer. of Chicago Press, 1941.
 25. MILLER, A. W. Confirmation of ergic structure in man, by *R*-technique analysis of objectively measured dynamic traits. Unpublished M.A. thesis, Univer. of Illinois, 1951.
 26. ROGERS, C. R. *Client-centered therapy*. New York: Houghton Mifflin, 1951.
 27. SAUNDERS, D. R. Practical methods in the direct factor analysis of psychological score matrices. Unpublished Ph.D. dissertation, Univer. of Illinois, 1950.
 28. SPENCE, K. W. The postulates and methods of behaviorism. *Psychol. Rev.*, 1948, 55, 67-69.
 29. STEPHENSON, W. Correlating persons instead of tests. *Charact. & Pers.*, 1935, 4, 17-24.
 30. STEPHENSON, W. The inverted factor technique. *Brit. J. Psychol.*, 1936, 26, 344-361.
 31. STEPHENSON, W. The foundations of psychometry: Four factor systems. *Psychometrika*, 1936, 1, 195-209.
 32. STERN, W. *Differentielle psychologie*. Leipzig. Barth, 1911.
 33. THOMSON, G. H. *The factorial analysis of human ability*. London: Univer. of London Press, 1948.
 34. THURSTONE, L. L. Second order factors. *Psychometrika*, 1944, 9, 71-100.
 35. THURSTONE, L. L. The effects of selection in factor analysis. *Psychometrika*, 1945, 10, 165-198.
 36. THURSTONE, L. L. *Multiple-factor analysis*. Chicago: Univer. of Chicago Press, 1947.
 37. WENIG, P. The relative roles of various ego defense mechanism operations in tests of misperception. Unpublished M.A. thesis, Univer. of Illinois, 1951.
 38. WILLIAMS, H. V. M. A *P*-technique study of personality factors in the psychosomatic areas. Unpublished Ph.D. dissertation, Univer. of Illinois, 1949.

Received March 10, 1952.

SOME GRECO-LATIN ANALYSIS OF VARIANCE DESIGNS FOR LEARNING STUDIES^{1,2}

E. JAMES ARCHER

Northwestern University³

In verbal learning studies having no more than about six conditions, it is frequently desirable to have the same *S* serve under all conditions. This procedure has been commonly described as having each *S* serve as his own control. To avoid a systematic bias because of a particular order, the conditions are usually arranged in a latin square. Of course, the use of such a design is defensible only when the interaction of the literal variable and rows, or literal variable and columns, will not be significant. However, this absence of interaction will occur with greater than rare frequency (see McNemar, 8). Briefly, in a latin square: (a) there are as many columns and rows as there are conditions, (b) a particular condition appears once and only once in each row and each column, and (c) each row of conditions has a different permutation of orders. The analysis of variance for the latin square design has been described by Grant (4). Since each *S* serves in all conditions, the rows usually correspond to different *S*s. Therefore, in a nonreplicated design there are as many *S*s as there are rows. To increase the sample size there are two alternative procedures for replicating, depending upon the hypotheses under examination. One may simply repeat the same square and have a test for order of conditions or, if order is of little interest, one may use permutations of the original latin square. A design such as the latter reduces the probability of a large systematic bias due to a particular latin square. These problems of replication have been discussed by Edwards (2, 3).

There is still another factor to consider. In most learning studies, *S* does not learn the same material time after time under all conditions. Rather, a different task, e.g., a different list of words, is learned under each condition. These lists may be learned by all *S*s in the same order,

¹ The writer wishes to express his appreciation to Professor David A. Grant for his reading and very helpful criticisms of this paper.

² The writer is indebted to Dr. Benton J. Underwood for his helpful suggestions in the writing of this paper. The analyses herein described were originally developed for the treatment of verbal learning data obtained under Contract N7onr-45008, Project NR 154-057, between Northwestern University and the Office of Naval Research. This project is under the direction of Dr. Underwood. The writer is also grateful to Dr. John W. Cotton for his critical reading of this paper.

³ Now at the University of Wisconsin.

in which case list differences would be confounded with a practice effect yet to be described, or these lists also may be presented according to another "latin" square independent of the square for conditions. Such a design is commonly called a greco-latin square. In the above example Latin letters would correspond to conditions and Greek letters would correspond to lists. The analysis of variance for the greco-latin design also has been described by Grant (4).

There are two kinds of practice effects, both of which are testable with an analysis of variance of repeated measurements. The first of these is the increase in performance made by *S* on successive trials when learning a *single* task. This we might designate as a specific effect of practice—specific for the single task being learned. The analysis of variance of repeated measurements for this specific practice effect has been dealt with by others (1, 6, 7). The second type of practice effect is observed as a change in *S*'s performance in the learning of several successive tasks. Here the effect of practice on several tasks belonging to the same class, e.g., lists of words, is more general than that which appears in the learning of a single task. This second case we might designate then as a general effect of practice—general for the class of tasks being learned.

It is this latter, or general, effect of practice which is testable in the proposed analyses. Since a different task is learned under a different condition, as outlined above, a systematic change in performance as a function of the ordinal sequence of combinations of conditions and tasks will be due to a general effect of practice. If, for example, the dependent variable were the number of trials required to learn the task to a given criterion, this general effect of practice would be designated a learning-how-to-learn effect.

To this point we have considered only a four-classification design: rows (*Ss*), columns (order of tasks), Latin letters (conditions), and Greek letters (learning materials). It is the purpose of this paper to describe analyses of variance of repeated measurements for higher orders of classification.⁴

⁴ In a recent article (9) the suggestion was made that an analysis of variance of repeated measurements on the same subject serving under different conditions was inappropriate. The evidence presented was for a special case, however. The *Ss* manipulated three variations of a Müller-Lyer illusion device and were not given knowledge of their results. Since there was considerable differential transfer from one condition to the other, it is true that a repeated measurements design was inappropriate. The acquisition of a persistent perceptual bias leading to differential transfer is unlikely in the usual verbal learning experiment.

A FIVE-CLASSIFICATION ANALYSIS WITH GRECO-LATIN SQUARES ACCOUNTING FOR TWO OF THE CLASSIFICATIONS

Description of a Hypothetical Experiment

In this design repeated measurements, e.g., "days" or "trials," account for one order of classification and define this method as a repeated measurements analysis of variance. A second classification is orthogonal to that of repeated measurements. Examples of this latter classification may be an ordered variable such as hours of food deprivation, or an unordered variable such as different types of rest interval activities in a distributed practice learning problem. This second classification contains a subclass, rows, which corresponds to different *Ss*. The data within this two-classification analysis may be further classified according to one or more greco-latin squares.

For these greco-latin squares, the Latin letters would account for a variable such as length of rest interval, whereas the Greek letters would account for different lists of words to be learned. Depending upon the nature of the problem, we may replicate the same greco-latin square, and consequently analyze a variance resulting from order, or we may permute the original greco-latin and reduce the probability of a systematically biased combination of Latins, Greeks, and orders. Generally we are concerned with the differences among conditions and not their order of presentation. Therefore, the discussion which follows will concern independently drawn or permuted greco-latin squares rather than a single greco-latin replicated. The reader may alter the analysis for a single greco-latin replicated following Edwards (2) if he wishes to test for differences among orders.

In Table 1 the schema is presented for this particular design. Six independent 4×4 greco-latin squares are presented. The numbers were taken from a table of random numbers (5), skipping zero entries.⁵

In this hypothetical experiment the numbers could indicate number of trials to learn a list of words to some criterion. The columns correspond to the successive days each *S* served in the experiment. The rows contain the four scores for each *S*. The blocks of three squares correspond to a particular kind of rest interval activity. The Latin letters indicate different lengths of rest intervals. The Greek letters indicate different lists of words to be learned.

⁵ Although this analysis has been used successfully on real data, a hypothetical design is presented here for brevity. Some verbal learning studies under the direction of Dr. B. J. Underwood involve as many as 108 *Ss* classified in three sets of 12 independently drawn 3×3 greco-latins. Presenting the original data would only unnecessarily increase the bulk of this discussion.

TABLE 1
RANDOM NUMBERS PRESENTED IN THE GENERAL FORM OF A FIVE-CLASSIFICATION*
ANALYSIS WITH PERMUTED GRECO-LATIN SQUARES ACCOUNTING
FOR TWO CLASSIFICATIONS

Block	Group	Greco-Latin Square				Ss	Days				Sum
							I	II	III	IV	
X	(a)	B ₁	D ₂	A ₃	C ₄	1	2	3	1	5	11
		A ₂	C ₃	B ₄	D ₁	2	7	5	4	8	24
		D ₄	B ₁	C ₂	A ₃	3	5	9	1	8	23
		C ₁	A ₄	D ₃	B ₂	4	3	7	2	5	17
		Sum					17	24	8	26	75
	(b)	A ₁	D ₂	C ₃	B ₄	5	9	9	3	7	28
		C ₁	B ₂	A ₃	D ₄	6	6	2	4	9	21
		B ₃	C ₄	D ₁	A ₂	7	7	8	8	6	29
		D ₃	A ₄	B ₁	C ₂	8	9	5	2	3	19
		Sum					31	24	17	25	97
	(c)	D ₁	B ₂	A ₃	C ₄	9	3	6	7	4	20
		A ₄	C ₁	D ₂	B ₃	10	4	5	5	4	18
		B ₄	D ₃	C ₂	A ₁	11	5	5	5	4	19
		C ₂	A ₁	B ₃	D ₄	12	3	1	5	3	12
		Sum					15	17	22	15	69
	(a)	A ₁	B ₂	D ₃	C ₄	13	7	4	3	5	19
		B ₄	A ₃	C ₁	D ₂	14	8	9	6	1	24
		C ₃	D ₄	B ₁	A ₂	15	1	1	8	3	13
		D ₁	C ₂	A ₄	B ₃	16	7	4	4	1	16
		Sum					23	18	21	10	72
	(b)	A ₁	D ₂	B ₃	C ₄	17	9	6	2	2	19
		D ₃	A ₄	C ₁	B ₂	18	1	3	4	3	11
		C ₄	B ₁	D ₂	A ₃	19	1	4	8	7	20
		B ₂	C ₃	A ₁	D ₄	20	1	6	3	5	15
		Sum					12	19	17	17	65
	(c)	A ₂	B ₁	C ₃	D ₄	21	3	2	4	4	13
		D ₁	C ₂	B ₄	A ₃	22	3	6	2	2	13
		B ₃	A ₄	D ₃	C ₁	23	3	5	5	1	14
		C ₁	D ₄	A ₁	B ₂	24	3	2	2	1	8
		Sum					12	15	13	8	48

* For the purposes of the two five-classification analyses the sixth classification of Group will be ignored. This classification will be used when the data are analyzed according to a six-classification procedure.

In this particular design each *S* serves in only one row. Consequently, 24 *S*s have participated in this hypothetical experiment. In addition to isolating the sums of squares of each of these main effects, we may expect three simple interactions: (a) Columns \times Blocks, (b) Latins

\times Blocks, and (c) Greeks \times Blocks. Ordinarily in a latin or greco-latin square design we would not expect to have an interaction for the Latin or Greek letters with a main effect, but in this design we have an amalgamation of the usual greco-latin design into a factorial design; hence the interactions.

General Form of Analysis

The general equation for this analysis is as follows:

$$\begin{aligned} \sum (X_i - \bar{X})^2 = & n_C n_{S/B} \sum (\bar{X}_B - \bar{X})^2 + n_S \sum (\bar{X}_C - \bar{X})^2 \\ & + n_S \sum (\bar{X}_L - \bar{X})^2 + n_S \sum (\bar{X}_\Gamma - \bar{X})^2 \\ & + \sum^B n_C \sum^S (\bar{X}_S - \bar{X})^2 + n_{S/B} \sum (\bar{X}_{BC} - \bar{X}_B \\ & - \bar{X}_C + \bar{X})^2 + n_{S/B} \sum (\bar{X}_{BL} - \bar{X}_B - \bar{X}_L + \bar{X})^2 \quad [1] \\ & + n_{S/B} \sum (\bar{X}_{B\Gamma} - \bar{X}_B - \bar{X}_\Gamma + \bar{X})^2 \\ & + \sum^B \sum^S (X_i - \bar{X}_C - \bar{X}_L - \bar{X}_\Gamma - \bar{X}_S + 3\bar{X}_B)^2 \end{aligned}$$

where X_i = any single score

\bar{X} = Grand mean

\bar{X}_B = Block mean

\bar{X}_C = Column mean

\bar{X}_L = Latin mean

\bar{X}_S = Subject mean

\bar{X}_Γ = Greek mean

\bar{X}_{BC} = cell mean based on Block-Column classification

\bar{X}_{BL} = cell mean based on Block-Latin classification

$\bar{X}_{B\Gamma}$ = cell mean based on Block-Greek classification

n_C = number of Columns

$n_{S/B}$ = number of Ss within block

n_S = total number of Ss

n_B = number of Blocks

N = total number of scores

It is to be understood that when a summation sign with a limiting superscript precedes another summation sign with a limiting superscript, the first superscript governs the limit of the second and all others within a given quantity, e.g.,

$$\sum^G \left[\sum^S (\sum X)^2 / n_C \right] \text{ means } \sum^G \left[\sum^{S/G} (\sum X) / n_C \right].$$

The degrees of freedom for the general case would be analyzed as in Table 2. The df for the Pooled $S_s \times$ Columns is shown as $n_B(n_{S/B}n_C - n_{S/B} - n_C - n_L - n_T + 3)$. This appears as if the value within the parentheses is multiplied by the number of blocks. Although this procedure is admissible when there are an equal number of S_s within all blocks, it is to be understood that the df associated with these $S_s \times$ Columns is actually obtained by summing for all blocks. This is of theoretical significance only with respect to df but of practical significance with respect to the corresponding sums of squares.

TABLE 2
DISTRIBUTION OF DEGREES OF FREEDOM FOR BASIC FIVE-CLASSIFICATION DESIGN

Source of Variation	df
Blocks	$n_B - 1$
S_s /Blocks	$n_B(n_{S/B} - 1)$
Columns	$n_C - 1$
Columns \times Blocks	$(n_C - 1)(n_B - 1)$
Latins	$n_L - 1$
Latins \times Blocks	$(n_L - 1)(n_B - 1)$
Greeks	$n_T - 1$
Greeks \times Blocks	$(n_T - 1)(n_B - 1)$
Pooled $S_s \times$ Columns/Blocks	$n_B(n_{S/B}n_C - n_{S/B} - n_C - n_L - n_T + 3)$
Total	$(n_{S/B})(n_B)(n_C) - 1$

The computational formulae are as follows:

- (1) Correction factor = $C = (\sum X)^2/N$
- (2) Total $SS = \sum X^2 - C$
- (3) Blocks $SS = \sum^B (\sum X)^2/(n_{S/B})(n_C) - C$
- (4) S_s /Blocks $SS = \sum^B \left[\sum^S (\sum X)^2/n_C - (\sum X)^2/(n_{S/B})(n_C) \right]$
- (5) Columns $SS = \sum^C (\sum X)^2/n_S - C$
- (6) Latins $SS = \sum^L (\sum X)^2/n_S - C$
- (7) Greeks $SS = \sum^T (\sum X)^2/n_S - C$
- (8) Columns \times Blocks $SS = \sum^{CB} (\sum X)^2/n_{S/B} - \sum^C (\sum X)^2/n_S - \sum^B (\sum X)^2/(n_{S/B})(n_C) + C$

$$(9) \text{ Latins} \times \text{Blocks } SS = \sum^{LB} (\sum X)^2 / n_{S/B} - \sum^L (\sum X)^2 / n_S \\ - \sum^B (\sum X)^2 / (n_{S/B})(n_C) + C$$

$$(10) \text{ Greeks} \times \text{Blocks } SS = \sum^{LB} (\sum X)^2 / n_{S/B} - \sum^L (\sum X)^2 / n_S \\ - \sum^B (\sum X)^2 / (n_{S/B})(n_C) + C$$

$$(11) \text{ Pooled } Ss \times \text{Columns/Blocks } SS$$

$$= \sum^B \left[\sum X^2 - \sum^C (\sum X)^2 / n_{S/B} - \sum^L (\sum X)^2 / n_{S/B} \right. \\ \left. - \sum^L (\sum X)^2 / n_{S/B} - \sum^S (\sum X)^2 / n_C + 3(\sum X)^2 / (n_{S/B})(n_C) \right]$$

Analysis of Variance of Hypothetical Data: Basic Five-Classification Design

Using the computational formulae given in the preceding section the total variance of the data from the hypothetical experiment shown in

TABLE 3
ANALYSIS OF VARIANCE OF DATA FROM HYPOTHETICAL EXPERIMENT:
BASIC FIVE-CLASSIFICATION DESIGN

Source of Variation	Sum of Squares	df	Mean Square	F
Independent Observations				
Blocks	32.667	1	32.667	5.462*
Pooled Ss/Blocks	136.458	22	6.203	
Total between subjects	169.125	23		
Correlated Observations				
Columns	9.375	3	3.125	—
Columns \times Blocks	25.750	3	8.583	1.435
Latins	20.875	3	6.958	1.240
Latins \times Blocks	8.583	3	2.861	—
Greeks	10.875	3	3.625	—
Greeks \times Blocks	3.916	3	1.305	—
Pooled Ss \times Columns/Blocks	303.126	54	5.613	
Total within subjects	382.500	72		
Total for experiment	551.625	95		

* For df 1 and 22, $p_{.05}F=4.30$; $p_{.01}F=7.94$.

Table 1 was analyzed. This analysis is shown in Table 3 in the format suggested by Edwards (3). The advantage of this arrangement of component values is that the appropriate error terms become immediately obvious.

In terms of the hypothetical variables outlined earlier, we would conclude that there was a significant difference between the two rest interval activities as indicated by an F of 5.462 for Blocks.⁶ The Pooled Ss/Blocks⁷ is the proper error term for testing the Block means' estimate of population variance. This is true only if the Ss were randomly assigned to the Blocks, however. Such an assumption is made here.

A test for individual differences could be made by dividing the Pooled Ss/Blocks variance estimate by the Pooled Ss \times Columns/Blocks.⁸

The proper error term for the remaining variance estimates, based as they are upon correlated observations, will be the Pooled Rows \times Columns/Blocks. According to our hypothetical assignment of variables to coordinates, we find that none of the main effects is significant. Also none of the interactions of these main effects and rest-interval activities, Blocks, is significant.

Limitations

As may be seen through the use of an experimental design such as outlined above, a great deal of information can be obtained even though only 24 Ss were involved. In this example it was possible to evaluate the effect of four main variables and three simple interactions. Such a return strongly supports Grant's prediction (4, p. 441) that the [greco]-latin square design will find wide application in psychology. Nevertheless this design has certain limitations which should be considered. Although each datum is governed by four classifications, it does not yield interactions between all main effects. This could be a serious drawback if, for example, we wanted to test the significance of the length of rest interval \times practice interaction. According to the above proposed design such an interaction, Columns \times Latins, would not be isolable, and, in fact, is assumed to be not significant. However, the difficulty could be partially circumvented by the judicious assignment

⁶ Theoretically an F ratio of this magnitude should occur but once in about 20 drawings of samples of random numbers of this size.

⁷ Logically, of course, the Pooled Ss/Blocks can be used as the error term only if homogeneity of variance is demonstrable. When there are only two cases involved a two-tail F test will be shorter than a Bartlett test.

⁸ Now the assumption of homogeneity of variance must be met for the Pooled Ss \times Columns/Blocks. If the numerator does not satisfy such an assumption, each contributing Block would be tested separately.

of variables to coordinates. Before the experiment was performed *E* could decide which interaction tests would be important and assign his variables accordingly. For example, if we wish to test the significance of a practice \times length of rest interval, we could assign "length of rest interval" to the coordinate occupied by "rest interval activities." We would lose the test for a rest interval activity \times practice interaction, but we would still have a test for the main effect.⁹

A FIVE-CLASSIFICATION ANALYSIS OF VARIANCE WITH GRECO-LATIN SQUARES ACCOUNTING FOR TWO CLASSIFICATIONS AND THE LEARNING TASKS DEFINING A COORDINATE

This design is essentially a variation of the one discussed previously, but because of its wide applicability it deserves individual consideration. In this design the "nature" of the learning task is dimensionalized, i.e., the learning tasks define points along one of the coordinates. For example, with reference to Table 1, the Columns (days), Rows (*Ss*), and Latins (lengths of rest intervals) remain the same but the Blocks represent different degrees of intratask similarity. Let us assume Block *X* has high intratask similarity and Block *Y* low intratask similarity. What this really means is that the four lists which are learned by *Ss* in Block *X* will be so constructed as to have high intralist similarity and the four lists which are learned by *Ss* in Block *Y* will have low intralist similarity. In brief the "nature" of the learning tasks (Greeks) defines the main coordinate (Blocks).

This variation introduces a few modifications into the analysis of the total variance, the associated *df*, and the choice of error terms for *F* tests. Since different learning materials are learned by the *Ss* in each block, a significant difference between blocks would mean that the materials were truly different, or more precisely, that the "nature" of differences between blocks influenced the measure of learning. Such a finding could be of considerable importance for some hypotheses. However, a significant difference between material classified by Blocks would not be equivalent to a difference between materials classified by Greeks. In fact, following the first analysis, a difference between Greeks would be confounded by a difference between Blocks. The solution to this dilemma is to test for differences among Greeks within each Block. There will be as many *F* tests for Greeks as there are Blocks. A significant *F* ratio would mean the materials within (and defining) a Block were different with respect to the dependent variable measured. This

⁹ This manipulation would give us four blocks and the number of rest interval activities would have to be increased to be classified as a Latin or Greek.

variation in analysis alters equation (1) by eliminating a Greek \times Block interaction and accounts for Greek variation in terms of deviation from Block mean rather than Grand mean. The altered equation is as follows:

$$\begin{aligned} \sum (X_i - \bar{X})^2 &= n_C n_{S/B} \sum (\bar{X}_B - \bar{X})^2 + n_S \sum (\bar{X}_C - \bar{X})^2 \\ &+ n_S \sum (\bar{X}_L - \bar{X})^2 + \sum^B n_S \sum (\bar{X}_T - \bar{X}_B)^2 \\ &+ \sum^B n_C \sum^S (\bar{X}_S - \bar{X}_B)^2 + n_{S/B} \sum (\bar{X}_{CB} - \bar{X}_C \\ &- \bar{X}_B + \bar{X})^2 + n_{S/B} \sum (\bar{X}_{LB} - \bar{X}_L - \bar{X}_B + \bar{X})^2 \\ &+ \sum^B \sum^S (X_i - \bar{X}_C - \bar{X}_L - \bar{X} - \bar{X}_S + 3\bar{X}_B)^2. \end{aligned} \quad [2]$$

The breakdown for *df* is also slightly altered and is shown in Table 4.

TABLE 4
DISTRIBUTION OF DEGREES OF FREEDOM FOR VARIATION OF
FIVE-CLASSIFICATION DESIGN

Source of Variation	<i>df</i>
Blocks	$n_B - 1$
Pooled Ss/Blocks	$n_B(n_{S/B} - 1)$
Columns	$n_C - 1$
Columns \times Blocks	$(n_C - 1)(n_B - 1)$
Latins	$n_L - 1$
Latins \times Blocks	$(n_L - 1)(n_B - 1)$
Pooled Greeks/Blocks	$n_B(n_T - 1)$
Pooled Ss \times Columns/Blocks	$n_B(n_{S/B}n_C - n_{S/B} - n_C - n_L - n_T + 3)$
Total	$(n_{S/B})(n_C)(n_B) - 1$

The computational formulae remain the same as for the previous analysis except that:

Pooled Greeks

$$SS = \sum^B \left[\sum^T (\sum X)^2 / n_{S/B} - (\sum X)^2 / n_C(n_{S/B}) \right]$$

replaces formulae (7) and (10).

Analysis of Hypothetical Data: A Variation with Greeks Defining a Co-ordinate

The significant change in re-analyzing the data presented in Table 1

is that the Greek \times Block interaction is absent and more F tests are in order. The analysis of variance is presented in Table 5.¹⁰ All sources of variance are tested and interpreted as before except the Greeks/Blocks, each of which is tested for significance by dividing its variance estimate by its corresponding $Ss \times$ Column/Block, e.g., Greeks/Block $_X$ divided

TABLE 5

ANALYSIS OF VARIANCE OF DATA FROM HYPOTHETICAL EXPERIMENT:
A VARIATION IN WHICH LEARNING TASK DEFINES A COORDINATE

Source of Variation	Sum of Squares	df	Mean Square	F
Independent Observations				
Blocks	32.667	1	32.667	5.462*
Pooled Ss /Blocks	136.458	22	6.203	
Total between subjects	169.125	23		
Correlated Observations				
Columns	9.385	3	3.125	—
Columns \times Blocks	25.750	3	8.583	1.435
Latins	20.875	3	6.958	1.240
Latins \times Blocks	8.583	3	2.861	—
Pooled Greeks/Blocks	14.791	6		
Greeks/Block X	6.562	3	2.187	—
Greeks/Block Y	8.229	3	2.743	—
Pooled $Ss \times$ Columns/Blocks	303.126	54	5.613	
$Ss \times$ Columns/Block X	133.230	27	4.934	
$Ss \times$ Columns/Block Y	169.896	27	6.292	
Total within subjects	382.500	72		
Total for experiment	551.625	95		

* For df 1 and 22, $p_{.05}F=4.30$; $p_{.01}F=7.94$.

by $Ss \times$ Columns/Block $_X=0.443$. In both cases in Table 5 the ratio was less than 1.000. This would be interpreted to mean, in our hypothetical experiment, that the four lists in each of the two blocks did not differ significantly one from the other in learning difficulty. That the two sets of four lists were of unequal difficulty is indicated by the F ratio of 5.462 for the Blocks divided by Pooled Ss /Blocks.

¹⁰ This analysis has also been applied to real data but for convenience and continuity of comparison the same hypothetical data will be used again.

A SIX-CLASSIFICATION ANALYSIS OF VARIANCE WITH GRECO-LATIN SQUARES ACCOUNTING FOR TWO CLASSIFICATIONS

This analysis is like the first except that another coordinate has been added. Such an addition considerably increases the complexity of the analysis and in one respect also reduces its precision somewhat. The same hypothetical data given in Table 1 will be used in this example. Now, however, the classification of Group will be considered. These Groups are designated by the letters *a*, *b*, and *c* in Table 1. The hierarchy is: Blocks contain Groups, Groups contain Squares, Squares contain Rows, and Rows contain scores, each of which lies in a Column. Finally the conditions under which each score is obtained are governed by a pair of letters, one Latin and one Greek.

Although this added coordinate increases our information from the data, we sacrifice some precision in that one variance estimate is based upon a pool of three triple interactions. At present these triple interactions are not isolable. The Pooled Triple Interactions are determined by subtraction. This is not too serious a disadvantage since if one of the three interactions is significant and if the other two are not significantly restricted, the Pooled Triple Interaction divided by the Pooled

TABLE 6

DISTRIBUTION OF DEGREES OF FREEDOM FOR BASIC SIX-CLASSIFICATION DESIGN

Source of Variation	df
Blocks	$n_B - 1$
Groups	$n_G - 1$
Blocks \times Groups	$(n_B - 1)(n_G - 1)$
Pooled <i>S</i> _s /Groups/Blocks	$(n_{S/G} - 1)(n_G)(n_B)$
Columns	$n_C - 1$
Columns \times Blocks	$(n_C - 1)(n_B - 1)$
Columns \times Groups	$(n_C - 1)(n_G - 1)$
Latins	$n_L - 1$
Latins \times Blocks	$(n_L - 1)(n_B - 1)$
Latins \times Groups	$(n_L - 1)(n_G - 1)$
Greeks	$n_\Gamma - 1$
Greeks \times Blocks	$(n_\Gamma - 1)(n_B - 1)$
Greeks \times Groups	$(n_\Gamma - 1)(n_G - 1)$
Columns \times Blocks \times Groups	$(n_C - 1)(n_B - 1)(n_G - 1)$
Latins \times Blocks \times Groups	$(n_L - 1)(n_B - 1)(n_G - 1)$
Greeks \times Blocks \times Groups	$(n_\Gamma - 1)(n_B - 1)(n_G - 1)$
Pooled <i>S</i> _s \times Columns/Groups/Blocks	$n_B n_G (n_{S/G} n_C - n_{S/G} - n_C - n_L - n_\Gamma + 3)$
Total	$(n_{S/G})(n_C)(n_G)(n_B) - 1$

$S_s \times \text{Columns/Groups}$ will indicate this significance. Deciding which of the triple interactions is significant is even less precise. For the present a rational analysis must be used. Although these disadvantages appear almost to invalidate this design, it should be pointed out that a significant and important triple interaction is infrequent. Furthermore, in almost all cases the triple interaction involving the learning materials (Greeks) will be insignificant and unimportant since an effort usually will be made to attain uniformity in tasks.

General Form of Analysis

The distribution of the df for this design is shown in Table 6. Again, as in the five-classification analysis, it is to be noted that the Pooled $S_s \times \text{Columns/Groups/Blocks}$ is actually obtained by summing these interactions for Groups and Blocks and not simply by multiplying by $n_B n_G$ as indicated. For the df this is of theoretical significance, but for the sums of squares this concept is of practical importance.

The computational formulae for this six-classification analysis is as follows:

$$(1) \text{ Correction factor} = C = (\sum X)^2/N$$

$$(2) \text{ Blocks } SS = \sum^B (\sum X)^2/(n_{S/B})(n_C) - C$$

$$(3) \text{ Groups } SS = \sum^G (\sum X)^2/n_{S/G}(n_C)(n_B) - C$$

$$(4) \text{ Blocks} \times \text{Groups } SS = \sum^{BG} (\sum X)^2/n_{S/G}(n_C) \\ - \sum^B (\sum X)^2/(n_{S/B})(n_C) \\ - \sum^G (\sum X)^2/(n_{S/G}(n_C)(n_B) +$$

$$(5) S_s/\text{Groups/Blocks } SS$$

$$= \sum^B \sum^G \left[\sum^S (\sum X)^2/(n_C) - (\sum X)^2/(n_{S/G}(n_C)) \right]$$

$$(6) \text{ Columns } SS = \sum^C (\sum X)^2/n_S - C$$

$$(7) \text{ Columns} \times \text{Blocks } SS = \sum^{CB} (\sum X)^2/n_{S/B} - \sum^C (\sum X)^2/n_S \\ - \sum^B (\sum X)^2/(n_{S/B})(n_C) + C$$

- $$\begin{aligned}
 (8) \text{ Columns} \times \text{Groups } SS &= \sum^{CG} (\sum X)^2 / (n_{S/G})(n_B) - \sum^G (\sum X)^2 / n_S \\
 &\quad - \sum^G (\sum X)^2 / n_{S/G} (n_C)(n_B) + C \\
 (9) \text{ Latins } SS &= \sum^L (\sum X)^2 / n_S - C \\
 (10) \text{ Latins} \times \text{Blocks } SS &= \sum^{LB} (\sum X)^2 / n_{S/B} - \sum^L (\sum X)^2 / n_S \\
 &\quad - \sum^B (\sum X)^2 / (n_{S/B})(n_C) + C \\
 (11) \text{ Latins} \times \text{Groups } SS &= \sum^{LG} (\sum X)^2 / (n_{S/G})(n_B) - \sum^L (\sum X)^2 / n_S \\
 &\quad - \sum^G (\sum X)^2 / (n_{S/G})(n_C)(n_B) + C \\
 (12) \text{ Greeks } SS &= \sum^F (\sum X)^2 / n_S - C \\
 (13) \text{ Greeks} \times \text{Blocks } SS &= \sum^{FB} (\sum X)^2 / n_{S/B} - \sum^F (\sum X)^2 / n_S \\
 &\quad - \sum^B (\sum X)^2 / n_{S/B} (n_C) + C \\
 (14) \text{ Greeks} \times \text{Groups } SS &= \sum^{FG} (\sum X)^2 / (n_{S/G})(n_B) - \sum^F (\sum X)^2 / n_S \\
 &\quad - \sum^G (\sum X)^2 / (n_{S/G})(n_C)(n_B) + C \\
 (15) \text{ Pooled } Ss \times \text{Columns/Groups/Blocks } SS \\
 &= \sum^B \sum^G \left[\sum X^2 - \sum^S (\sum X)^2 / n_{C/G} - \sum^C (\sum X)^2 / n_{S/G} \right. \\
 &\quad \left. - \sum^L (\sum X)^2 / n_{S/G} - \sum^F (\sum X)^2 / n_{S/G} + 3(\sum X)^2 / (n_{S/G})(n_C) \right] \\
 (16) \text{ Total } SS &= \sum X^2 - C \\
 (17) \text{ Pooled Triple Interaction} \\
 SS &= (16) - (2) + (3) + (4) + \dots + (15)
 \end{aligned}$$

Although this is a six-classification design there are no interactions beyond three triple interactions. If this Pooled Triple Interaction does not prove to be significantly greater than the Pooled $Ss \times \text{Columns/Groups/Blocks}$, these two SS may be combined and divided by the combined df 's and used as a more reliable estimate of error variance. Throughout these analyses it is assumed that before two or more vari-

ance estimates are "pooled," the proposed pool meets a test for homogeneity of variance.

Analysis of Variance of Hypothetical Data: Basic Six-Classification Design

If as in Table 1 we let Blocks correspond to two different kinds of rest-interval activities, Groups correspond to three levels of work rate for these activities and Columns, Latins, and Greeks correspond to days, lengths of rest intervals, and learning materials, respectively, we are ready to re-analyze the data according to the proposed six-classification analysis. This analysis is presented in Table 7. Since the Pooled

TABLE 7

ANALYSIS OF VARIANCE OF DATA FROM HYPOTHETICAL EXPERIMENT:
BASIC SIX-CLASSIFICATION DESIGN

Source of Variation	Sum of Squares	df	Mean Square	F
Independent Observations				
Blocks	32.667	1	32.667	6.515*
Groups	32.813	2	16.407	3.272
Blocks \times Groups	13.395	2	6.698	1.336
Pooled Ss/Groups/Rows	90.248	18	5.014	
Total between subjects	169.123	23		
Correlated Observations				
Columns	9.375	3	3.125	—
Columns \times Blocks	25.750	3	8.583	1.499
Columns \times Groups	20.687	6	3.448	—
Latins	20.875	3	6.958	1.216
Latins \times Blocks	8.583	3	2.861	—
Latins \times Groups	60.937	6	10.156	1.774
Greeks	10.875	3	3.625	—
Greeks \times Blocks	3.916	3	1.305	—
Greeks \times Groups	15.437	6	2.573	—
Pooled Triple Interactions:				
Columns \times Blocks \times Groups	122.311	18	5.724	36
Latins \times Blocks \times Groups				
Greeks \times Blocks \times Groups				
Pooled Ss \times Columns/Groups/Blocks	83.756	18		
Total within subjects	382.500	72		
Total for experiment	551.625	95		

* For df 1 and 22, $p_{.05}F=4.30$; $p_{.01}F=7.94$.

Triple Interaction was not significantly greater than the Pooled $S_s \times \text{Columns/Groups/Blocks}$ ($F = 1.460$, 2.22 required for $p_{.05}$) these two variance estimates have been combined and used as an "error" term for testing the Columns, Latins, and Greeks and their interactions. As in the previous analyses, only the Block means are significantly different.

The interpretations and rationale of the tests remain the same as before. It is understood of course that the use of one greco-latin square per Group is only the minimum. Any multiple number of squares may be used as required by the reliability desired and other features of the experimental design. If one wished, he could derive a variation of this design with Greeks defining Blocks as previously described.

Limitations and Applications

The chief limitation of this design is the loss of precision in isolating the sums of squares for each of the triple interactions. This could be a serious deterrent in some studies. In the usual learning studies it is unlikely that the triple interactions will be of theoretical significance. In such cases a six-classification design could be very useful since much information can be gleaned from relatively little data.

Both the five- and six-classification analyses and their variations have wide applicability, assuming the nonsignificance of nonisolable interactions. In addition to those already discussed, examples of variables which could be tested by using the basic design in which the same materials are learned in all blocks are: methods of material presentation (audio, visual, audio and visual), rates of material presentation, levels of frustration induction, levels of motivation, e.g., ego-involvement studies, hours of sleep deprivation. Other variables which could be tested by using the variation design in which materials define blocks are: levels of meaningfulness, levels of affectivity, different types of presentation (paired associates versus verbal discrimination).

SUMMARY

Two basic multiple-classification analyses of variance of repeated measurements incorporating greco-latin squares were described in detail. A useful variation of one of the designs was also described. In each case these analyses were applied to 96 random numbers which were arranged in the appropriate experimental designs under discussion. The applicability and limitations of the methods were described.

REFERENCES

1. ALEXANDER, H. W. A general test for trend. *Psychol. Bull.*, 1946, **43**, 533-557.
2. EDWARDS, A. L. Homogeneity of variance and the Latin square design. *Psychol. Bull.*, 1950, **47**, 118-129.
3. EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.
4. GRANT, D. A. The latin square principle in the design and analysis of psychological experiments. *Psychol. Bull.*, 1948, **45**, 427-442.
5. KENDALL, M. G., & SMITH, B. B. Randomness and random sampling numbers. *J. roy. statist. Soc.*, 1938, **101**, 147-166.
6. KOGAN, L. S. Analysis of variance—repeated measurements. *Psychol. Bull.*, 1948, **45**, 131-143.
7. LINDQUIST, E. F. Goodness of fit of trend curves and significance of trend differences. *Psychometrika*, 1947, **12**, 65-78.
8. MCNEMAR, Q. On the use of latin squares in psychology. *Psychol. Bull.*, 1951, **48**, 398-401.
9. SMITH, B. B. On some difficulties encountered in the use of factorial designs and analysis of variance with psychological experiments. *Brit. J. Psychol.*, 1951, **42**, 250-268.

Received January 2, 1952.

A NOTE ON PROFILE SIMILARITY

HAROLD WEBSTER

University of Kentucky

The article by Osgood and Suci (2) on a measure of profile similarity deserves comment. The need for such measures is certainly great, as the authors point out, and their treatment of the problem is interesting. This note is intended to suggest other approaches to the same problem.

It is true that most correlation coefficients, when used as measures of profile similarity, disregard the absolute differences between means of profiles. One exception, however, is intraclass r . Intraclass r between two parallel profiles decreases in value from +1.00 toward zero as the profiles are moved farther apart. It has the advantage of having a known standard error, but the disadvantage of possessing a restricted negative range (1).

The problem concerning the orthogonality of measurement variates should be made more explicit. It is true that if measurement variates are orthogonal, that is to say independent (and thus uncorrelated), then D has the value given by the authors. In any random sampling problem, however, it is unlikely that five measures such as those in their Fig. 1 would have zero intercorrelations. The problem of allowing for the effects of such intercorrelations has been a central one in multivariate analysis. Unfortunately the complete solution for comparing profiles is indeterminate when each profile represents only one person and hence provides only a single degree of freedom. If each profile should represent means of persons in a small group, then D could be obtained from

$$D^2 = \sum_i^k \sum_j^k a^{ij} d_i d_j \quad (i, j = 1, \dots, k),$$

where a^{ij} is the matrix inverse to a_{ij} , the measurement variate dispersion matrix for the entire sample, and d_i, d_j are differences between means for the two groups. This is Mahalanobis' D^2 , a statistic which not only allows for the effects of the intercorrelations among the k variates but also may be tested for significance by an F ratio (3).

Rao (4) has given a transformation for the original measurement variates which makes them mutually independent. This transformation is lengthy if there are more than six or eight variates, but has the advantage of avoiding the even lengthier inversion of the dispersion

matrix which is ordinarily necessary for computing the D^2 . With this transformation Mahalanobis' D becomes the D used by Osgood and Suci.

REFERENCES

1. FISHER, R. A. *Statistical methods for research workers*. London: Oliver and Boyd, 1948.
2. OSGOOD, C. E., & SUCI, G. J. A measure of relation determined by both mean difference and profile information. *Psychol. Bull.*, 1952, 49, 251-262.
3. RAO, C. R. Tests of significance in multivariate analysis. *Biometrika*, 1948, 35, 58-79.
4. RAO, C. R. On a transformation useful in multivariate computations. In P. C. Mahalanobis, D. N. Majumdar, et al., *Anthropometric survey of the United Provinces, 1941: A statistical survey*. *Sankhyā*, 1949, 9, 89-324. Appendix 5, pp. 251-253.

Received June 21, 1952.

ON THE RELIABILITY OF THE LEADERLESS GROUP DISCUSSION TECHNIQUE

BORIS SEMEONOFF

University of Edinburgh

Ansbacher's recent survey *The History of the Leaderless Group Discussion Technique* (1) quotes only one reference to work on reliability of the technique, that of Bell and French (2). Evaluation was in terms of mutual ranking "in order of preference for discussion leader." Although the relevance of this criterion is not altogether clear, the work did at any rate yield an average correlation of mean ranks for each candidate, within the various groups in which he participated, of 0.75. This must be accepted as a positive result, although regarded as a reliability coefficient it is not outstandingly high.

The object of this note is to call attention to an alternative method of evaluation, based not on mutual rating, but on an objective assessment of participation.

The underlying principle is similar to that of the sociometric method of Moreno and others. Essentially it is based on the recognition of two types of leadership in group discussion, which may be designated, respectively, as *active* and *passive*. The former may be measured by the number of times a member of the group speaks, and the latter by the number of times he is spoken to. It would perhaps be feasible to evolve a formula combining these two variables which might give an optimum estimate of effectiveness in the discussion situation, although it will of course be recognized that effectiveness cannot be measured in quantitative terms alone. However, the derivation of such a formula is not relevant to the present purpose.

Work along these lines was carried out by the author in connection with a selection procedure similar to that described by the OSS Staff (4). No observer participated in the discussion, or entered the circle in which the candidates sat, except to explain the procedure at the outset. The discussion lasted about forty minutes, divided into two approximately equal periods, with a different topic for each period. The progress of the discussion was recorded in such a way as to show not only the number of times each member of the group spoke or was spoken to, but also the sequence of the various contributions to the discussion. This record, however, was kept purely for research purposes, and was not used in rating effectiveness in the group discussion situation; such rating was carried out according to principles common to most of the writers referred to by Ansbacher.

From the record four "raw" measures may be obtained, and re-

garded as two measures each of "active" and "passive" leadership in the group discussion situation, as defined above. These are referred to in what follows as AI, AII, PI, and PII, the respective topics being designated as I and II.

If high correlations are found between AI and AII and between PI and PII, we may conclude that reliability of these measures has been demonstrated, and consequently that a meaningful formula for "leadership" in this situation may also yield a reliable measure. In other words, there would appear to be a relatively stable structure within the group when engaged in this particular activity. If, on the other hand, correlations between AI and PI and between AII and PII tend to be higher than the reliability coefficients for A and P, it would suggest that any hierarchical structure within the group during the discussion of a particular topic was more closely related to such factors as knowledge (or even only knowledgeability) in the relevant field, than to any permanent structure within the group as a group.

Detailed results have not been published, and are not now available, but the second of the possibilities outlined above was that which emerged in nearly every case. Reliability was thus seen to be low, seldom rising above the level of a correlation of the order of 0.4. This finding would also appear to cast serious doubt on the validity for practical purposes of a discussion based on a single randomly chosen topic, since knowledge or interest would appear to be a dominant factor. It would also appear to be an argument in favor of what may be described as the "object-stimulus" method of group discussion latterly adopted by War Office Selection Boards in Great Britain. This method makes use of a freely developing discussion based on an informal conversational opening similar to one which might be used among a group of strangers in a railway carriage. Reference to this method is made by Harris (3), but it is hoped that a further discussion of its rationale will be published later. Whether such a device leads to a more reliable discussion situation has not, so far as the writer knows, been investigated.

REFERENCES

1. ANSBACHER, H. L. The history of the leaderless group discussion technique. *Psychol. Bull.*, 1951, 48, 383-391.
2. BELL, B., & FRENCH, R. L. Consistency of individual leadership position in small groups of varying membership. *J. abnorm. soc. Psychol.*, 1950, 45, 764-767.
3. HARRIS, H. *The group approach to leadership testing*. London: Routledge & Kegan Paul, 1949.
4. OSS ASSESSMENT STAFF. *Assessment of men*. New York: Rinehart, 1948.

Received March 10, 1952.

GENETICS OF SCHIZOPHRENIA: A REJOINDER

NICHOLAS PASTORE

Queens College

Since space limitations do not permit a detailed answer to Hurst's reply (1) to my review (3) of Kallmann's *Genetics of Schizophrenia* (2) I shall select a few of the major points for discussion. Most of the inaccuracies of understanding and misquotations which Hurst introduces in his reply will have to be ignored.

1. Hurst asserts that Haldane and Hogben acclaimed Kallmann's work. In response to an inquiry by the writer Haldane wrote: "I regret that I possess no copy of the monograph to which you refer, and have no recollection of having ever read it, or commented on it any way, favourably or otherwise. Mr. Hurst may, of course, attribute this blank in my mind to senile dementia, Freudian repression, or some other failing on my part. From what little I have read on the genetics of schizophrenia it appears to me that its genetical investigation presents grave difficulties. In view of its variable age of onset and uncertain diagnosis I am sceptical as to any statements made about its genetics."

2. Despite Hurst's claim to the contrary, Kallmann did include the "doubtful" schizophrenics in the "definite" schizophrenic category in many of his analyses. My statement concerning the 16.4 per cent figure (and other figures as well) as including the "doubtfuls" was based on computations in accordance with Kallmann's procedure. Since it can be arithmetically shown that the "doubtfuls" were included in Kallmann's theoretical discussions my criticism of Kallmann's demonstration of a "gene-coupling" between the hereditary dispositions of tuberculosis and schizophrenia still holds. (The relevant calculations can be made available to the interested reader.)

3. In contradiction to my statement in the review Hurst asserts that Kallmann "invariably" included numbers as well as percentages in the subgroups. Hurst is in error. Kallmann's Table 10 does not state the numbers for *any* category with the exception of the second column entitled "absolute." Kallmann's Table 38 does not state the numbers for any cell. Other examples exist.

4. Hurst alleges that I failed to understand Kallmann's Table 58 because I did not grasp the significance between "expectancy" and "net" figures. This table deals with the health outcome of those children who have parents both of whom are schizophrenic. According to Kallmann, the probability that such offspring will develop schizophrenia is 68.1 per cent and the probability that such offspring will develop schizoidia is 45.7 per cent. I criticized these probabilities because their sum is significantly larger than unity whereas the sum of independent probabilities, mathematically speaking, would not exceed unity (Kallmann

presents no computation of the probability that the offspring will remain healthy). Therefore, I inferred the presence of "overlapping" categories. This may or may not be the case. The absurd result of a sum of apparently mutually exclusive (and not even exhaustive) probabilities which exceeds unity must be explained. Hurst fails to offer an explanation.

5. In my original review I demonstrated a numerical contradiction between the 108 secondary cases enumerated in Chapter 5 of Kallmann's book and the enumeration of the same cases in Tables 34-37. Hurst alleges that I am in error because of my apparent failure to distinguish between "expectancy" and "net" figures. However, the distinction between the two types of figures is not involved and the discrepancy is a real one, as the reader can verify for himself by summing the "absolute" numbers in the relevant tables. This point is adequately discussed in the original review.

6. In criticism of Kallmann's Table 10, I stated that "overlapping categories" were involved (3, p. 294). Hurst states that this is not the case (1, p. 407). Let us examine this table. The subcolumn headings entitled "schizophrenia in one parent" and the other entitled "schizophrenia in one parent and other abnormalities in the second parent" do overlap. The first phrase asserts nothing about the qualities of the other parent and can very well subsume the second phrase. I admit that this might be too literal an interpretation of the subcolumn headings. However, there is another "overlap" within this table. A given proband could have a schizophrenic parent and also a schizophrenic grandparent and/or a schizophrenic aunt or uncle as well. Thus this particular proband could be placed in three separate columns (inflating the total percentages). The percentages in the last row of the table need not at all refer to different probands. Kallmann provides no information which would enable the reader to ascertain the degree of overlap.

7. In view of Hurst's comments on the reliability of the clinical records used by Kallmann (1, p. 405f) I wish to clarify one of my statements in this regard. My point referred to the possible change in observing and recording clinical symptoms as a result of Kraepelin's nosological contributions. One of Kraepelin's criteria for distinguishing between dementia praecox and manic-depressive psychosis was the presence of progressive deterioration in the patient. The introduction of this criterion may have led psychiatrists to look for and record those symptoms in the patient which seemed to be in line with the criterion. In pre-Kraepelinian days such a directed outlook was formally lacking and the determination of diagnoses could have been affected. Therefore an error of "indeterminate magnitude" is involved in the original clinical records of the probands—the initial set of observations with which Kallmann began his investigation.

It is indeed remarkable that Hurst is incorrect in *all* his allegations. I could only discuss a few of his objections in this brief note. The judgment I expressed in the review that the "Kallmann investigation . . . supplies no reliable information for assessing the genetic basis of schizophrenia" (3, p. 302) is emphasized by the weakness of Hurst's reply. One final remark—in the review I did not deny the possible genetic etiology of schizophrenia.

REFERENCES

1. HURST, L. A. Genetics of schizophrenia: Reply to Pastore. *Psychol. Bull.*, 1951, **48**, 402-412.
2. KALLMANN, F. J. *The genetics of schizophrenia*. New York: J. J. Augustin, 1938.
3. PASTORE, N. The genetics of schizophrenia: A special review. *Psychol. Bull.*, 1949, **46**, 285-302.

Received February 18, 1952.

THE GENETICS OF SCHIZOPHRENIA: FURTHER REJOINDER TO PASTORE

LEWIS A. HURST

Alexandra Institution, Maitland, Cape Town, South Africa

In my original reply to Pastore my allusion to Haldane's and Hogben's view of Kallmann's work applied quite clearly to their reaction to his paper read at the Seventh International Congress of Genetics, Edinburgh, 1939 (my reference 5), the Proceedings of which I summarized for *Mental Hygiene* (my reference 1). Kallman's Edinburgh address concerned his American twin-family study to date, and not his earlier German study which constitutes the subject matter of his monograph *The Genetics of Schizophrenia* (my reference 3). There is thus no contradiction between my claim and Haldane's quoted assertion concerning the book.

So much for section 1 of Pastore's rejoinder. In replying to the other sections, I do not propose reiterating tabular and statistical details as our conflicting claims are now fairly and squarely before our readers for their arbitrament. I shall merely rehearse general principles and elaborate and clarify my original reply where this appears necessary.

The first part of Pastore's contention in section 2 has already been dealt with to my satisfaction. Regarding the latter portion concerning the genetics of tuberculosis and the hypothesis of gene-coupling between it and schizophrenia, it was not perhaps made sufficiently clear in my original reply that in the 1943 articles written in collaboration with Reisner (my references 10 and 11), Kallmann quite explicitly, on the

basis of new evidence, modified his views as to *details of*, although *not the fact of* heredito-constitutional mechanisms in tuberculosis. The magnitude and significance of Kallmann and Reisner's contribution are forcibly brought home by the tribute of Barbara S. Burks on the occasion of the 1943 presentation of the findings.¹ One may add that the work of Kallmann's tuberculosis unit, exploring more detailed clinical and genetic aspects, has gone on unremittingly since 1943, and that a paper from Kallmann, Reisner, and Planansky is expected in the near future.

Section 3 of Pastore's rejoinder is the most veritable casuistry. I used the word "invariably" regarding numbers in subgroups for a frame of reference defined by Pastore's original criticisms, that is, where they are statistically relevant. The same considerations apply to cell entries. Pastore cannot claim that his play upon words here affects any scientific conclusion.

Sections 4 and 5 point, in my opinion, to a confusion between net and expectancy figures persisting in Pastore's mind. The fallacy of Pastore's criticism in his section 4 directed against table 58 in Kallmann's monograph, on the grounds of the sum of the independent probabilities of schizophrenia and schizoidia in the offspring of double schizophrenic parentage exceeding unity, is exposed by the following considerations. First, the table reveals quite clearly that of the 55 offspring of such unions, 32 only were schizophrenic or schizoid (16 of each). Then it must be borne in mind that the corrected rates of reference for schizophrenia and schizoidia here are 23.5 and 35 respectively, differing because of the different age distribution (apparent from the table) with reference to the manifestation period of the psychosis. To demand that the sum of these disparately derived expectancy figures should not exceed 100 per cent is ideologically unsound. An understanding of Weinberg's Abridged Method would have saved Pastore from this pitfall.

In section 6 Pastore objects to what he terms overlapping categories on the grounds of their inflating morbidity figures. He explains this by the example of a subject who may be entered in 3 columns in virtue of having not only one schizophrenic parent but also a schizophrenic grandparent and a schizophrenic aunt or uncle as well, and so counting as 3 instead of 1 when columns are totalled. What a quibble this is becomes apparent from scrutinizing Table 10 in Kallmann's monograph to which he specifically refers. It is found to consist of a series of 5

¹ BURKS, BARBARA S. Comment on paper by Kallmann and Reisner. *Amer. Rev. Tuberculosis*, 47, 572.

tables separated by double lines (not columns of a single table) and there is no intention whatsoever of adding the totals. In cases where a subject genuinely plays a double or multiple role, Weinberg's proband and sibling method is specifically devised to correct the resulting bias. This method is fully described and frequently utilized in Kallmann's monograph.

One last word—I would ask the reader to turn his attention away from the minutiae of genetic interpretation to general outlines, which the hairsplitting of the present debate may have obscured. There is altogether too much of this sort of thing abroad today. Witness Bronson Price's attempt on the basis of anomalies in twin development to discredit twin studies in general and those of Kallmann in particular—although viewing these phenomena in perspective shows that they are of such rare occurrence as to render them incapable of significantly affecting the concordance rates, their differences and similarities, in representative samples of adequate size such as are employed by Kallmann. The long tradition of twin studies which springs from Galton is not likely to collapse in the face of such ill-founded attacks—witness Gedda's encyclopaedic work on twins (1950) and his recently founded twin journal.

Received June 23, 1952.

BOOK REVIEWS

HUMPHREY, GEORGE. *Thinking: An introduction to its experimental psychology*. New York: Wiley, 1951. Pp. xi+331. \$4.50.

Five adjectives summarize the feelings I have about this volume: important, difficult, honest, scholarly, and historical. A defense of these adjectives should give a prospective reader some notion of the kind of book Professor Humphrey has written.

Important. This adjective is easy. The subject itself is important and no good summary has been available. This book is the best secondary source for research on thinking that you can buy today.

Difficult. Most of the book is difficult reading and some of it could even be called dull. This adjective gets less applicable as the text proceeds into more modern and more familiar territory. The most difficult chapters are those dealing with the Würzburg group and with Selz. These men found conscious processes that defied clear description. They invented or adapted German words to name them. Even in German these names connote more than they denote. Their concepts were painted on an intricate, highly elaborated, theoretical background that is completely inadequate and unfamiliar today. Translating from German to English and from 1910 to 1950 is a tremendous task in itself. It is not Humphrey's fault that the result is hard to read. The worst charge against him is that he devoted half the book to this translation.

Honest. The reading would have been easier if the author had been less honest in his reporting. The white lie, the begged question, the suggestive analogy, the over-generalization, and all the other dishonest devices that give respite to a weary reader are no part of this book. Each worker is reviewed systematically and usually is quoted at some length. Criticisms are given in the man's own terms and in the light of what he tried to do. No one is glibly denounced because he is dead and will not reply.

Scholarly. Patient scholarship shows through every page. Most psychologists are content with secondary sources and trust Boring, Woodworth, or Murphy for their opinions. We may scan some original sources and even read a few. But it is clear that Humphrey has studied the originals carefully until he understands what they say and what they do not say. The only place his diligent scholarship falters is in the decade from 1940 to 1950.

Historical. The organization is largely historical, beginning with associationism, adding the imageless thoughts and determining tendencies from Würzburg, then moving to Gestalt theories, motor theories, language, and generalization. Almost three-fourths of the references are dated prior to 1935. Rare and inaccessible sources are profusely quoted to preserve them for the future.

Because of his concern to keep the record straight and to deal honestly with each contributor, it is rather difficult to see what Humphrey's own position is. His earlier book, *Directed Thinking*, 1948, gives a clearer picture of Humphrey's general stand than does the present one. In many respects, the present volume is a teacher's manual to accompany the simpler version of 1948.

At the end of the book Humphrey outlines the present position of the experimental psychology of thinking. Fifty years of research are summarized in 16 statements. Thinking occurs when an organism meets, recognizes, and solves a problem. A problem is a situation that holds an organism from its goal. Thinking combines features of the problem which were originally discrete. It involves past experience. The form and method of the ingression of the past into the present is under dispute. Trial-and-error occurs. Motive is an aspect of thinking. Thinking is directed. The Würzburg group believed thought to be free from sensory content. They neglected the image which is a form of organization. Gestalt theorists have stressed productive as opposed to reproductive thinking. Thinking may be accompanied by changes in muscular tonus. Language cannot be equated with thinking. Generalization is a constant response to an invariable feature in a variable context. Images, action, speech, and concepts organize responses to a problem. Meaning is an artificial problem. At the rate of one such sentence for every three years of work, we have a long time to wait before we understand thinking. Perhaps the man who summarizes the first century will have at least one quantitative function worth drawing as a graph.

As the author says, "Fifty years' experiment on the psychology of thinking or reasoning have not brought us very far, but they have at least shown the kind of road which must be traversed." The prospect is not encouraging, but future travellers will surely agree that Professor Humphrey helped to show the way.

GEORGE A. MILLER.

Massachusetts Institute of Technology.

KEMPTHORNE, O. *The design and analysis of experiments*. New York: Wiley, 1952. Pp. xi+631. \$8.50.

This book is a worthy addition to Wiley's distinguished and growing list of titles in statistics. Although it is of little use to the beginning student, it can be profitably studied and used by psychological researchers at three different levels of statistical sophistication.

1. The experimenter who understands the meaning of least-squares methods and has some acquaintance with analysis of variance techniques will find a systematic and readable account of the tests with which he is familiar and can achieve a better appreciation of the standard tests and their possible variants even if he does not follow the mathematical details.

2. The student who commands only the elementary calculus can follow the general argument of the book (matrix arguments are usually set aside in separate sections) and will find a unified underlying treatment that should increase his understanding of the relations among the various tests.

3. For the reader who can handle matrix reasoning in addition to the calculus the book provides a satisfactory and eminently readable discussion of many

of the theorems underlying the statistics we use and the relations of these theorems to the interpretation of experimental data.

The reviewer particularly recommends an examination of Kempthorne's book to anyone who has attempted a systematic increase of his statistical prowess and found Fisher's works too elliptical and such books as Wilks' *Mathematical Statistics* too formidable mathematically.

The first five chapters are devoted to a brief discussion of scientific method with special consideration of the role of statistical interpretation and to a presentation and discussion of fundamental statistical ideas. More attention is given to problems of statistical estimation than is usually the case in the books on applied statistics widely used by psychologists. The teacher of advanced statistics should find these early chapters useful. The reviewer believes he can attest to the pedagogical soundness of the presentation given by Kempthorne, having worked out a similar approach (though neither so concise nor so complete as Kempthorne's) for use in his own classes.

In the later chapters a detailed discussion of an extremely large number of designs is given. These chapters should be of great value to the individual researcher and of perhaps even more value to those overworked members of departmental staffs whose duty it is to supervise the statistical aspects of graduate student research. The table of contents is detailed and descriptive so that it is easy to find information on any specific design.

In sum, this seems an excellent book for use either as text or as reference.

C. J. BURKE.

Indiana University.

JAHODA, MARIE, DEUTSCH, MORTON, COOK, STUART W., and others. *Research methods in social relations: With especial reference to prejudice. Part one: Basic processes. Part two: Selected techniques.* (2 vols.) New York: Dryden Press, 1951. Pp. x+759. \$3.75 each vol., \$6.00 the set.

These two volumes are intentionally different from each other in organization and level of discourse. The first of the volumes is the result of a cooperative effort among Jahoda, Deutsch, and Cook, and attempts to present an introductory, integrated account of "the considerations which enter into every step of the research process" (p. v) generally as well as those aspects of it specifically related to the area of social relations, with the illustrative material drawn from the area of prejudice where possible. The second volume is composed of a set of short chapters by various authors on methodological problems selected with the intent of supplementing in more technical detail some of the issues discussed in *Part One*. The audience at which the books are directed

is broad: lay persons who are to act upon the findings of social scientists, students who are preparing to do social research, and social scientists who have not specialized in the field of social relations. These, then, are the knowledge areas and people aimed at by the works' authors. In evaluating how well they have hit their targets, neither the white bull's-eye disk nor the red, miss flag is appropriate.

Certainly, the approach is a laudable one; rather than the usual cold, abstract, machine-like picture of the research process, it is portrayed here as being actual behavior of real people and, therefore, beset by the same troubles as any like kind of social behavior. Certainly, the emphasis upon the necessity for coherent, explicit prior planning from hunch and hypothesis to data analysis and interpretation is commendable. Worth while, too, is the attention given to often neglected aspects of research such as fiscal and personnel administration. In general, then, where the books deal with what to do, how to do it, and when to do it, they are solidly based and make a real contribution. Nevertheless, some things about them are disturbing. Inasmuch as the volumes are so different in organization, these points will be taken up separately for each.

Part One. Unevenness in the level of discourse exists. Anyone who has to be told in Chapter Two that research is not a willy-nilly data collection affair is seldom ready to have Solomon's extension of control group design and the terms "analysis of variance" and "analysis of covariance" tossed at him in the next chapter. Examples of similar character may be found throughout the book and one suspects that sometimes the authors were considering their lay audience and at others the student or social scientist group. This flaw is not major in this volume; the authors' attitude and argument regarding values in research are, however.

Speaking to this point, it is easy to be misunderstood. Therefore, first, let it be asserted that few psychologists today would deny that values, personal and social, play a large part in their selection of a research problem area. Nor would it be denied by most that the social scientist is not removed from his social, institutional and cultural milieus; his rôle is that of the citizen-scientist. But it is a far cry from those assertions to the position that the scientist must act to produce social change, and yet this is the way the authors' argument runs. They appear to deny this by saying, "There is no valid scientific argument to compel or forbid [the social scientist] to encourage an [action] agency to apply his results" (p. 320). Yet this recognition is more verbal than real, since continual and heavy stress is laid upon research done for the purpose of promoting socially desirable objectives where "desirable" can only be defined by the researcher. And it is easy for the reader to infer that the authors as scientists (not citizens or even citizen-scientists) know scientifically what is good for society. This kind of interpenetration of personal values with data and action based upon those data is dangerous: it can lead to the espousal and support of an intellectual dictatorship potentially as bad (my valuation) as any other dictatorship; quieter, perhaps, but just as authoritarian.

As to the book's difficulties with theory, they may stem from the attitudes

expressed about "action research." The following quotation appears to sum up the authors' own felt conflict between citizen and scientist: "One of the severest obstacles to an acceleration of the process of theory formation is the fact that those engaged in field research of the sort we have discussed in this book have not enough time to devote to it" (p. 335). And, even though a great deal of space is devoted to theory in the book, it remains a nebulous topic. As the book is read, one question keeps recurring: "All right. Theory is good and hypotheses are necessary. How are they constructed and used?" Then, too, one encounters statements like this: "If according to [the measuring] instruments, the prediction is not borne out, and if the value of the theory is beyond doubt (and this is a big "if" in the social sciences), [the investigator] will conclude that the instruments did not measure what they were designed to measure" (p. 110). Such a statement is naive. Is any theory ever beyond doubt even if it is not in the area of social science? Shouldn't a theory specify the empirical operations by means of which it may be tested?

Part Two. The most notable flaw in this volume of eleven chapters, each by different authors, is its extreme unevenness. Of course, unevenness is expected in a collection such as this, but the range here goes from an exhortatory chapter by Wormser and Selltiz on community self-surveys to a technical chapter on sample design by McCarthy. More editorial supervision would seem to have been indicated. (This last comment might also be applied to the organization of the two volumes taken together; there is a great deal of repetition of material.)

Value judgments and general problems of theory play a minor part in this volume; these are how- and what- and when-to-do chapters for the most part. Several of them are excellent references for the research worker. Notable among these are Kornhauser's guide to questionnaire and interview schedule construction, Proctor and Loomis' review of methods of analyzing sociometric data, and Stouffer's revision of Chapter I, Volume IV of *The American Soldier* series on scaling theory.

By way of summary, let it be said that considerable space has been spent on the negative aspects of the two books under consideration. This fact may leave the wrong impression. When these books deal with theory and values in science generally, it is easy to quarrel with them or find their approach inadequate, but when they deal with the so-called practical aspects of research, it must be said that they attain their goal of pulling together on an introductory level a large amount of valuable but otherwise scattered material on research in social relations.

J. C. GILCHRIST.

University of Wisconsin.

TILTON, J. W. *An educational psychology of learning*. New York: Macmillan, 1951. Pp. vii+248. \$3.50.

This book treats the conditions of learning, particularly those that seem to have significance for education, from the viewpoint of field psychology. It should find a place as a basic or supplementary text in undergraduate courses. Its usefulness as such is enhanced by the inclusion of several chapters on educational measurement and the nature

of intelligence that are largely independent of the author's systematic position. The book as a whole is too elementary for the graduate student of education, except for the nonspecialist in educational psychology.

Considerable space is given to the applications of field psychology to education. As has been true of others writing from this standpoint, the author appears to have been most successful in his handling of attention and perception as factors in learning, and least successful with motivation, particularly the acquisition of secondary motives. The subject of efficiency in learning also appears to be difficult to treat from the standpoint of field psychology.

In a work of this kind, it is always a question as to the extent to which educational ideas and practices are actually demanded by a particular system, or the extent to which they are as compatible with one system as another. In this respect, Tilton's book shows up rather well. In the first place, it is conservative in its claims, and in the second, it brings forth some ideas that stem from its systematic position. An example is the envisagement of the educative process as the "planned introduction of novelty in the experiences of the learner." Here the emphasis upon new patterning and new organizations of experience and reorganization demanded by new data is in the Gestalt tradition.

The book is scholarly and reserved. It is singularly free of extravagant and irritating claims that characterize much of current educational writing espousing field psychology. For example, the author does not see atomism as a necessary attribute of behaviorism or *S-R* theory, nor does he see trial-and-error and insightful learning as being mutually exclusive processes. Chiefly, he thinks a broader perspective is required since so much in human learning is poorly described in terms of reinforcement. Perhaps it will occur to many readers of his book that the reinforcement and field psychologists may sometimes work at different levels of description.

J. B. STROUD.

State University of Iowa.

FILM REVIEWS¹

PRONKO, N. H., & SNYDER, F. W. *Vision with spatial inversion*. 18 min.; silent; black and white. State College, Pa.: Psychological Cinema Register, 1951. Rental, \$2.00 per day; sale, \$37.00.

This film shows the changes that occurred in certain motor performances of a man during the course of an experiment in which inverting lenses were worn continuously for 30 days. The tasks which were systematically explored were: walking a complicated and very irregular chalk line, mirror-tracing of a star-shaped path, Minnesota Manipulation Test, Purdue Pegboard Test, card-sorting. For these the subject is shown performing *before* wearing the lenses, during his *first* wearing of the lenses, after wearing the lenses *24 days*, at the end of the experiment (*30 days*), and immediately after *removing* the lenses. Aside from stating when the performance occurred, there are few verbal titles and no interpretation.

The apparent objective of the film is to demonstrate that although inverting lenses cause a tremendous upset in the subject's motor behavior at first, it is possible in time to learn to react nearly as quickly and accurately with them on as without them. The qualitative evidence of this learning process is excellent, but it is regrettable that learning curves are not given to show the quantitative changes induced as a function of time.

The film would be appropriate in an elementary psychology course, as a supplement to the discussion of visual perception, presumably in connection with a consideration of the role learning plays. With considerable supplementary reading it might be helpful in an advanced course on perception or when dealing with the field of perception in an experimental course.

The chief criticism the present reviewer would make of the film is that it fails to *show* the psychological experience of the subject; it is stated that the lenses reverse the field of vision from top to bottom and from right to left, but this is no more than a textbook would state. There seems to be no reason why the film might not incorporate, for example, a picture of a room in the laboratory as it looks normally and as it appears when wearing the lenses.

A second omission is the failure to demonstrate the difference in learning tasks like card-sorting, which is definitely visually controlled, and the Minnesota Manipulation Test, which is almost exclusively kinesthetic. Presumably the learning curves for these two tasks were quite different, and undoubtedly they also showed a marked difference during the first lens trials, as compared to performance before putting on the lenses, but this is not brought out.

¹ *Editor's Note:* These reviews of films were prepared under the auspices of the Committee on Audio-Visual Aids of the American Psychological Association. Dr. A. A. Lumsdaine was chairman of the Committee at the time the reviews were written.

On the whole, this is a good job. It gives a feeling for an important perceptual problem and develops it clearly and understandably; the photography is good. Any theoretical interpretations will have to come from the instructor using the film; whether this is a constructive or a destructive criticism will, obviously, be a function of one's personal bias.

DOROTHEA E. JOHANNSEN.

Tufts College.

MILLER, NEAL E., & HUNT, GARDNER L. (assisted by Douglas Lawrence and Reign Hadsell). *Motivation and reward in learning.* 15 min.; sound; black and white. State College, Pa.: Psychological Cinema Register, 1948. Rental, \$2.50 per day; sale, \$60.00.

This film illustrates the significance of motivation and reward in the learning of albino rats. Its general thesis is that motivation activates the subject sufficiently to produce a wide range of responses. If one of these responses leads to a reward, i.e., a reduction of drive, the tendency to repeat that response will be increased. The wide range of irrelevant activities is gradually narrowed as the tendency to make the rewarded response is increased, until the subject is directly and efficiently performing the response that leads to the reward.

This thesis is demonstrated by observing the behavior of several rats in a modified version of a Skinner-box. The film starts with comparing a hungry and a satiated albino rat. The hungry rat is shown actively exploring the glass-front box. His responses are quick and varied. In the course of these activities he presses a bar which releases a pellet of food in the food cup. Gradually, his activities become confined to the vicinity of the bar and food cup. Eventually he learns to press the bar and eat the food with a minimum of wasted action.

A satiated rat, who has been placed in an identical apparatus, is first shown resting complacently. A different kind of motivation and reward is illustrated when a shock is presented to this satiated rat through the grid floor. His activity increases sharply and he quickly learns to depress the bar which, in this instance, ends the shock.

The general point is then made that *any* response which the subject is capable of making can be learned if it is followed by a reward. We are briefly shown several different rats in the same apparatus who have learned, respectively, to turn a wheel, bite a rubber tube, or fight with each other when these responses led to the cessation of shock. The film ends with a good summary of its main points.

It can be judged from the above synopsis that this film presents a simplified version of the role of motivation and reward in learning as conceived by S-R reinforcement theorists. This job it does very well. Observation of the rats in the film illustrates the point of view and

procedure in a way that would be difficult to achieve through lectures or reading. It is perhaps best compared with a laboratory period or demonstration. Over these procedures it has the advantage (at least for students of elementary psychology) of eliminating the fumbling and presenting only significant details.

It is the opinion of this reviewer that the film would be a very useful adjunct to a lecture on the role of motivation and reward in learning for the introductory course and for undergraduate experimental psychology classes, particularly if the lecturer finds the S-R reinforcement approach palatable. But since it is a good demonstration of instrumental conditioning, it is likely that teachers with different theoretical inclinations can also find a place for it in their courses.

TRACY S. KENDLER.

New York University.

HAYES, K. J., & HAYES, C. *Vocalization and speech in chimpanzees*. 12 min.; sound; black and white. State College, Pa.: Psychological Cinema Register, 1950. Rental, \$2.25 per day; sale, \$50.00.

The film stresses two concepts. First, the chimpanzee will consistently utilize specific vocal sounds under specific forms of emotional stimulation. These specific vocal sounds are typical of the species in general and do not result from social imitation. The subject of this demonstration was isolated from others of the same species from birth. Sound samples of some of these typical vocalizations are presented:

The food bark during the period of anticipation of food.

The soft cry during low-level apprehension.

The high scream during high-level apprehension.

Laughter during tickling.

Second, the chimpanzee can be trained to vocalize certain sounds not naturally used; food objects served as motivation for the training and continuation of this behavior. Examples of these trained vocalizations are presented. These vocalizations have a rough imitative quality to the human sounds of mama, papa, and cup. The unnatural, trained sounds appear distinctively less smooth and effortless than normal spontaneous sounds. Examples of trained mouth movements are also supplied.

If the objectives of the film were to present clear-cut experimental results, these objectives have been impressively met. However, there is an absence of theoretical explanation, and of indication of the ramifications of the experimental results presented. Without such supporting interpretation, the film appears to have its greatest usefulness in advanced courses in general psychology, where theoretical aspects of the problems are known, or can be easily comprehended by the student when presented by the instructor.

The film appears to be average in photography, ingenuity of technique, and editing. The medium of the sound film has been excellently used to transmit the sounds of the chimpanzee voice.

EDWARD M. BENNETT.

Tufts College.

STONE, L. J. *When should grown-ups help?* 13 min., 16 min.; sound; black and white. New York: New York University Film Library. Rental, \$4.00 per day; sale, \$60.00.

STONE, L. J. *And then ice cream (children's meals).* 10 min., 16 min.; sound; black and white. New York: New York University Film Library. Rental, \$4.00 per day; sale, \$45.00.

These two films were produced at Vassar College in the series known as "Studies in Normal Personality Development." They are intended as class discussion material, for training of students in the techniques of observation with nursery-age children.

When Should Grown-Ups Help? presents four brief episodes in each of which a child is "helped" by an adult:

1. Andy is building a tall tower of blocks. In order to complete its top, he needs help. His teacher gives him the help. Then he is ready to start another tower-building project.

2. Penny is putting on her over-clothes. The adult (teacher or parent?) tries to help her, but Penny adamantly refuses, even though the adult keeps trying to help. Finally, Penny gets her clothing on without help.

3. Donald needs help in pounding a long nail into wood to join it with another piece of wood. He keeps hammering, but he makes no headway. He wants help but he gets none. He cannot complete his task.

4. Charlotte has caught her foot in the rope connecting her tricycle and cart. She twists and turns, while the teacher stands by without helping. Finally, Charlotte extricates herself.

And Then Ice Cream presents several cases of nursery-school children who are eating their school lunch. One eats slowly but methodically. Another rejects most of the meal. Ice cream dessert is considered an integral part of the meal, but it is reserved for the time *after* the food-eating period. Thus the ice cream dessert serves as a motivating factor in developing good eating habits.

These films may be considered most suitable for use with classes studying techniques of observation of nursery-school children. For such classes both the films present simple incidents which could stimulate much class discussion. In this connection, each film presents a set of discussion questions for consideration by the class. It is doubtful however, whether the films in their present form would be much suited for other purposes. They lack the wealth of clinical material which is found in many of the Vassar films, such as "This Is Robert" and "Frustration Play Techniques."

Photographically, the prints submitted for review appeared somewhat grainy, but this condition may be corrected in other prints. Technically, a device was employed which the present reviewer found most distracting. Whenever the passage of time was represented, a caption, "TIME," separated the sequences. This device may have been an attempt to substitute for a mechanical "fade in-fade out" or "lap-dissolve," but the net result was not satisfactory, giving a rather "jumpy" sequence instead of smooth transitions.

A major fault of each film is the procedure used for review of the film content. The producers have simply repeated the film content verbatim. While this method of review has some merit under certain circumstances (e.g., if no one is available to rewind the film), the same result could be achieved by showing the film over again. Rather than review by repetition, a more appropriate procedure might have included summarization, or interpretation of some of the behavior demonstrated.

ELIAS KATZ.

Sonoma State Home.

VERPLANCK, W. S. and associates. *Testing intelligence with the Stanford-Binet*. 18 min.; sound; black and white. State College, Pa.: Psychological Cinema Register, 1950. Rental, \$3.25 per day; sale, \$75.00.

This is a simple presentation of the method of determining IQ with the Revised Stanford-Binet Scale.

An examiner administers some of the subtests to different children. Mental age (MA) and IQ are computed for children of similar chronological age (CA). A brief glimpse is afforded of some uses of intellectual evaluation in teaching, guidance, and counseling.

Since it is not intended as an intensive training film on the techniques of administration of the S-B Scale, the film would be most useful in elementary psychology and education courses. It would be especially helpful in courses where students had relatively little opportunity for practice in administering and scoring this test.

While the photography and the commentator's talk throughout are quite professional, the recording of the examiner's voice seems somewhat fuzzy in the early part of the film.

ELIAS KATZ.

Sonoma State Home.

BENNETT, A. E., & MCKEEVER, L. G. *"Antabuse" in the treatment of alcoholism*. 17 min.; silent; color. State College, Pa.: Psychological Cinema Register, 1950. Rental, \$3.50 per day; sale, \$85.00.

Interest in the use of Antabuse in the treatment of alcoholism is widespread. Consequently, any portrayal of the manner of its use and its effects are to be welcomed. The film shows the reaction which occurs

when alcohol is drunk by a patient under treatment with Antabuse. By actually seeing what takes place in such circumstances, the viewer obtains an account of the reaction which would enable him to recognize it. In addition, information is given concerning the circumstances under which Antabuse therapy should and should not be undertaken. Roughly the latter half of the film follows a patient from the time of his admission to the hospital until his discharge. Some indication is given as to the place of Antabuse in the treatment armamentarium.

The film could be used successfully in a course in abnormal psychology or social problems. It could very well illustrate a lecture on the treatment of alcoholism. Although the film assumes some knowledge of medical terms, ignorance in this area is no handicap. Viewers should, however, have some orientation concerning alcoholism before seeing the film. The film might be shown to police officers who will have greater occasion to see the sufferer from an Antabuse reaction as use of the drug increases.

ALBERT D. ULLMAN.

Tufts College.

e. By
er ob-
ize it.
under
oughly
admis-
as to

ychol-
on the
edge of
ould,
seeing
greater
e drug

MAN.

B

I

I